Psychological Bulletin

HARRY HELSON, Editor
University of Texas

CONTENTS

Published Bimonthly by the
American Psychological Association

VOL. 58, NO. 3 MAY 1961

The *Psychological Bulletin* contains evaluative reviews of research literature and reviews of research methodology and instrumentation in psychology. This JOURNAL *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to the Editor, Harry Helson, Department of Psychology, University of Texas, Austin 12, Texas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50–60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

Reprints. Fifty free reprints are given to contributors of articles and notes.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D.C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin

and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A.

VOL. 58, NO. 3 MAY 1961

Psychological Bulletin

METHODOLOGY AND RESEARCH ON THE PROGNOSTIC USE OF PSYCHOLOGICAL TESTS

SAMUEL C. FULKERSON AND JOHN R. BARRY[1]
Western Psychiatric Institute and Clinic, University of Pittsburgh

There has not been a general review of the use of psychological tests in prognosis since Windle's review in 1952. At that time Windle concluded that, (*a*) it appeared to be some characteristic of the patient rather than the therapy given which determined the outcome of mental illness; (*b*) most studies in the area were difficult to interpret due to inadequate specification of one or more of the following: the sample characteristics, the treatment schedule, the criteria of improvement, and the degree of control imposed on variables influencing outcome; (*c*) the necessary step of cross-validation was usually omitted; and (*d*) personality tests, including the projective tests, had shown little promise in predicting outcome.

The purpose of the present article is to bring the review of the research on the prognostic use of tests up-to-date and to deal with some related methodological issues. The scope and organization will depart from that used by Windle. Firstly, the present review covers a wider range of criteria. Windle considered primarily the problem of predicting improvement. However there seems to be a complex of criteria which are closely related, logically and in practice, and so articles have been included dealing with a variety of criteria other than improvement. Secondly, the organization will differ from Windle's. He centered his review around individual tests, taking each test in turn and citing all prognostic studies where it had been used. The present paper is organized around the predictive problem rather than the individual test, since in practice the clinician wants to know how to come to a decision about a patient rather than what can be done with a given test. It is hoped that this emphasis will help to point up which questions are involved in the area of prognosis, and the relative attention each has received in research. And finally, the emphasis on decisions reflects an interest in decision theory (Luce & Raiffa, 1957), which has recently been suggested (Cronbach & Gleser, 1957) as a promising frame of reference from which to regard psychodiagnostic testing.

Windle included studies from as early as 1926 through 1951. The present review mainly covers the period from 1952 through June 1959. The coverage is more complete for those sections dealing explicitly with the prognostic use of tests than for the sections on methodological problems. Only the major psychological and psychiatric journals have been reviewed exhaustively.

[1] Acknowledgment is due the criticism and additions of Charles Windle and Joseph Zubin.

Methodology

The methodological difficulties in research on prognosis concern the researcher's decisions as to what samples he will use, what selection instruments he will apply to the sample, and what criteria seem most appropriate.

Sample Attributes

One of the primary methodological difficulties has been the definition of the sample. Psychiatric diagnosis appears to have been the predominant basis of sample definition in spite of the known unreliability of these categories. Attributes of the sample such as age, education, sex, or socioeconomic status are usually listed. However, little attention has been paid in most of the studies reviewed to achieving homogeneous samples or subsamples. Some investigators have worked with only one diagnostic group, mainly schizophrenics. Since schizophrenia is a diagnosis given to over 50% of unspecified functional psychotic disorders, the difference between the results from such studies and those using psychotics sampled at random are hard to determine.

The need for homogeneous samples is clearly pointed up by a consideration of the question of base rates. Meehl and Rosen (1955) said "a psychometric device, to be efficient, must make possible a greater number of correct decisions than could be made in terms of the base rate alone" (p. 194). Studies of base rate as a function of diagnostic category (Langfeldt, 1956; Pascal, Swensen, Feldman, Cole, & Bayard, 1953; Rennie, 1953) indicate wide fluctuation in outcome between categories. Examination of these base rates indicates that a sample of psychotic patients with a preponderance of manic-depressives would have a higher base rate of improvement (approximately 68%) than a sample consisting of schizophrenics (approximately 50%). A predictor, even though its actual validity was zero, could do much better predicting outcome in the first sample than in the second if the cutting point of the predictor was adjusted to take advantage of the percentage of improvement. The optimal chance percentage of correct prediction in the first sample (achieved by calling everyone improved) would be 68%, which is equal to the base rate. With a sample of any size this would differ significantly from 50%, the value likely to be designated as chance if one did not know the base rate. And if the relative effectiveness of a predictor in the two samples were tested it could appear, spuriously, that the predictor was 68% correct within one sample but only gave 50% correct prediction in the other. Thus, prognostic research designs which compare the results in the experimental group against statistical chance, or which compare two small groups that are not sufficiently matched on variables related to base rates, cannot result in useful information. Since effective handling of the problem of sample homogeneity is uncommon in the prognosis studies reviewed by Windle and ourselves, the generality of findings is low, or at best difficult to determine.

It has been assumed that homogeneous sampling represents an effective way of solving the problem of sample definition. However there is one danger. If the basis on which the homogeneity is established is highly related to the criterion variable, the variability of the criterion will be restricted. This can of course obscure a relationship that might exist between a predictor and criterion. It has been

tacitly assumed that adequate randomization is not easily achieved in prognosis research, considering the usual sample size and the biases of the clinical populations from which they are drawn; otherwise random sampling would be an efficient way to select, and thus operationally define, the sample.

Another difficulty with diagnosis as a basis for the definition of the samples is that it represents clinical judgments which are based upon an often uncertain weighting of situational and response variables. For instance, the diagnosis of depressive reaction typically requires a differentiation as to whether the affective response reflects anxiety or depression, and a decision concerning the degree to which the affective response is related to a currently stressful situation. Clearly clinicians can vary as to the relative emphasis they place on these variables; and, as several studies (Glass, Ryan, Lubin, Reddy, & Tucker, 1956; Gleser, Haddock, Starr, & Ulett, 1954) have shown, they do vary in their weighting procedures. Therefore, despite the convenience of using diagnosis as a sample-defining operation, it is weak in that the researcher loses some control over the basic stimulus and response elements upon which the judgment is based. Studies of the effects of these elements on test validity and on the efficiency of cutting scores are called for; this kind of research, frequent in personnel psychology, seldom appears in the clinical journals.

Tests

Here the primary difficulty has been to define that universe of test behavior related to outcome. The majority of studies cited in Windle's earlier review used standard tests, e.g., Rorschach, TAT, MMPI. It is likely that these tests tap only a small part of the response spectrum. With the welter available, it is still far from clear how many separate functions they sample, and no definitive taxonomy of tests exists. Zubin and his co-workers (Burdock, Sutton, & Zubin, 1958; Burdock & Zubin, 1956; Zubin, 1958, 1959) have proposed five broad categories of activity to which test behavior can be assigned: physiological, sensory, perceptual, psychomotor, and conceptual. Each of these five categories has been further subdivided into classes of stimuli and responses. For the most part, the prognostic measures which Zubin has selected to use within each category are simpler than such tests as the Rorschach, in the sense that they present fewer stimulus dimensions and require less elaborate and lengthy responses. Such systems of categorization may indicate a range of tests available, but within each category there is a degree of complexity which at this time is largely unknown. However, factor analyses have been carried out in the areas of perception (Thurstone, 1944), psychomotor tests (Fleishman & Hempel, 1954, 1956; Hempel & Fleishman, 1955; Seashore, Buxton, & McCollum, 1940), and cognition (Guilford, 1956, 1959). Such analyses afford at least a partial basis for rational test selection.

Since a number of studies (e.g., Conrad, 1954) indicate that severity of mental illness is a significant prognostic variable, researchers looking for simple tests for prognostic studies may find it of value to consider studies in the area of differential diagnosis. H. E. King (1954) was able to differentiate between chronic schizophrenics, subacute behavior disorders, and normals using psychomotor tasks; and Eysenck, Granger,

and Brengelmann (1957), with groups similar to those used by King, found a large number of both simple and complex perceptual tests which discriminated between their groups. Rabin and G. F. King (1958) reviewed studies dealing exclusively with schizophrenia, and concluded that "relatively high discriminatory power . . . has been obtained with simple experimental tasks. In many cases it has been as good as or better than that found with more complex tasks" (p. 253).

Criteria

There are three broad aspects of prognosis in mental illness: duration, course, and outcome. Studies predicting duration have used criteria such as length of hospital stay, the amount of time spent on the admitting or disturbed ward before transfer to a less disturbed ward (Gordon, Lindley, & May, 1957), and length of treatment.

Criteria involving the course of illness include measures of termination and relapse. In inpatient settings premature termination has been defined as leaving the hospital against medical advice; in outpatient settings it has been variously defined as not appearing for the initial interview after making an appointment, dropping out of therapy before some stipulated minimal number of contacts, or dropping out of therapy against the wishes of the therapist.

All criteria of improvement have been classified in this paper as measures of outcome. It could be argued that change over time is a measure of the course of illness, but this category has been reserved for specific qualitative aspects of change. Current criteria of improvement present the same difficulties in definition, and for this reason it is convenient to deal with them together, and to separate them from termination and relapse criteria. Because there is no universally agreed upon definition of the term "mental illness" (Jahoda, 1958; Scott, 1958), there has been a concomitant lack of clarity about how to measure its alleviation. Three sources of improvement criteria are common: (*a*) ratings of improvement made by the therapist, the patient, or other persons in contact with the patient such as relatives, professional staff other than therapist, or even fellow patients; (*b*) changes in objective measures of functioning such as physiological changes, or improvement in psychological test performance; and (*c*) follow-up data of a behavioral nature, such as whether patient is able to get and hold a job, to get or remain married, or, in whatever way, to resume a minimally independent social existence.

There have been several attempts to systematize these various outcome measures. An early breakdown of the separate areas of behavior which should be evaluated was made by Knight (1941). He suggested that therapists look for change in these five areas of adjustment: the disabling symptoms or problems, the interpersonal relations, the sexual adjustment, the productivity (i.e., the ability to work effectively and to utilize available energy), and the ability to handle stress. In Zubin's classification of tests, the ability to handle stress is viewed as a general parameter which might apply to the other four areas.

Barron (1953b) listed five similar criteria of improvement: (*a*) the patient feels better—indicated by introspective comments by the patient; (*b*) the patient relates better to others —requiring a follow-up at work, school, or home, and often based on

reports of members of the patient's social group; (*c*) the patient's symptoms clear up—as measured by psychiatric ratings of improvement at discharge, as well as indirectly by measures of duration, e.g., length of hospital stay and speed of transfer to minimum security wards; (*d*) the patient makes decisions in a health-tending direction; and (*e*) the patient's verbal behavior shows increased "insight."

A few other criteria have occasionally been proposed to supplement these. Winder (1957) has suggested changes in the adjustment of children of the patient, and Morse (1953) has proposed accessibility to psychotherapy. Reznikoff and Toomey (1959) list in detail a variety of attempts to provide a taxonomy of outcome criteria.

There are measurement problems in all of these approaches. Scott (1958) has pointed out several conceptual and methodological difficulties in the various definitions of mental health. His discussion can be applied to Barron's criteria of improvement in mental health: (*a*) apparent change in subjective feelings or symptomatology can be a function of change in environmental conditions or can be distorted by defense mechanisms; (*b*) difficulties in social relationships can be a function of the differing requirements of socioeconomic and cultural systems, and can change as the patient changes his community or his contacts in the community; (*c*) there can be disagreement over which is a health-tending direction, since value systems are frequently involved; and (*d*) changes in insight may be a function of the degree to which the patient is willing to conform to the theory and values of the therapist. It should be noted that these points need not be regarded as criticisms of the definitions. If, for instance, changes in subjective feelings are considered important in their own right, then changes in feelings, whether due to environment or defense mechanisms, are still of interest. However, when used as criteria, such changes are meant to reflect specific intra-individual changes that are independent of environmental or irrelevant personal factors. Despite this, most of the research in prognosis seems designed to demonstrate only that characteristics of the patient exist which relate to outcome, without controlling sufficiently for the above mentioned environmental and personal factors.

On a less general level, Parloff, Kelman, and Frank (1954) have listed several common sources of ambiguity in improvement criteria: (*a*) improvement is often treated as a unitary concept, but this may be erroneous; (*b*) the emphasis of the rater can interact with aspects of the treatment—for instance, symptoms typically disappear before insight occurs, so that a rater who requires signs of insight before he gives a rating of improvement will judge fewer patients to be improved than one who accepts symptom alleviation as improvement; and (*c*) improvement is likely to be overestimated, since patients fluctuate in behavior, and at any given time signs of improvement in one or more specific areas are likely to be present and thus overvalued by a judge being asked to make a global, subjective rating. Pascal and Zax (1956) criticize the usual gross "improved-unimproved" criterion on the grounds that it is not sufficiently tailored to the specific desired changes of the patient. They reject all nonbehavioral criteria of improvement, and essen-

tially appear to feel that symptom-change should be the primary criterion of improvement.

It would be valuable to know the factorial structure of the above course, duration, and outcome measures. While no study was found which attempted to do this, several reported intercorrelations between two or more prognostic criteria. These will be described separately for the kinds of criteria involved.

Correlations between outcome measures. Kelman and Parloff (1957) intercorrelated a number of measures, including ratings of comfort and self-awareness made by the patient, and social effectiveness ratings made by persons close to the patient as well as professional observers. The change in rating from pretherapy to 20 weeks after the initiation of therapy was determined. Only 1 of 21 intercorrelations between these measures of change was found to be significant. However, the correlations were based on an *N* of only 15, and the period of time was perhaps too short to expect more than minimal changes.

Storrow (1959, 1960) compared ratings of improvement made by therapists, patients, relatives of the patient, and a psychiatrist who had access only to abstracted material. Two related rating clusters were found: the patient's self-rating, the relative's rating, and the rating made by inexperienced therapists (third year medical students) formed one cluster; with the experienced therapist and the nontherapist psychiatrist forming the other. The correlations within clusters ranged from .61 to .79; between clusters, .32 to .57. These two clusters seemed to reflect primarily a dichotomy between patient and experienced therapist, since the relatives, and apparently the inexperienced therapists, gained their impression from hearing the patient's views of his progress, while the nontherapist psychiatrist obtained his knowledge from the file written by the therapist. Storrow had the ratings made separately for each of Knight's (1941) five areas, and the average intercorrelation between areas was approximately .60. Ellsworth and Clayton (1959) found that a measure of ward adjustment at discharge correlated significantly (.47) with a 3-month follow-up rating of community adjustment. However, amount of psychopathology at discharge had no relationship to the follow-up criterion. Their finding can be compared with the intercorrelation of .57 reported between two simultaneous ratings of adjustment made on different scales (Stilson, Mason, Gynther, & Gertz, 1958).

Patient expressions of positive and negative feelings have been used as evidence of improvement (see Auld & Murray, 1955, for a review of these measures). Barry (1950) found low but significant correlations between these so-called internal or feeling criteria and global judgments of improvement in adjustment. Rogers and Dymond (1954) have found that changes in patient self-ratings on *Q* sorts correlated with ratings and other criteria of improvement. In an analogous group research program Snyder (1953) reported that self-rating changes correlated significantly with judgments of improvement. The same results have been reported by Kalis and Bennet (1957). Taylor (1955) found that self-ratings (*Q* sorts) tend to become increasingly positive simply with the passing of time. This suggests that it is imperative to control for time in treatment in order to evaluate the actual extent of the relationship between self-ratings and other improvement criteria.

Correlations between duration and outcome. Ullman (1957) reported that a measure of length of hospital stay correlated .36 ($N=72$) with a measure of adequacy of interpersonal relationships (Palo Alto Group Therapy Scale), those rated most adequate after a period of group therapy being the ones with the shortest hospital stay. Pascal et al. (1953) found a correlation of .37 ($N=486$) between length of hospital stay and ratings of improvement made a year after discharge; again, the greater the improvement, the shorter the hospital stay.

A significant positive relationship has been frequently reported (Bailey, Warshaw, & Eichler, 1959; Myers & Auld, 1955; Seeman, 1954; Sullivan, Miller, & Smelser, 1958) in which greater length of psychotherapy in outpatient settings is accompanied by judgments of greater improvement. An interesting exception to this is the phenomenon called the "failure zone."

D. S. Cartwright (1955) found a grossly linear relationship between the number of psychotherapy sessions and success of outcome as noted by the therapists; but the mean success rating dropped sharply for those whose therapy lasted from 13 to 21 interviews. Cartwright was reporting on cases treated by nondirective techniques. Taylor (1956) validated this "failure zone" in a psychoanalytically oriented setting. Standal and van der Veen (1957) obtained the same drop in a counseling center sample. Vosburg (1958), in an examination of treatment charts, found evidence that from the fifteenth to twentieth hour was a period where outpatients tended to be preoccupied with their relationship with the therapist, suggesting that treatment which ended in this period might often be due to a desire on the part of either the patient or therapist to avoid the close, dependent relationship which was developing. Perhaps supplementing this, Ends and Page (1959) reported that the "flight into health" reaction occurred in group psychotherapy uniformly around the fourteenth session.

Correlation between duration and course. Crandall, Zubin, Mettler, and Logan (1954) found a significant relationship between the duration of initial hospitalization and rehospitalization; patients who stayed in the hospital a short time were most likely to still be out of the hospital on 1 to 4 year follow-up.

To summarize these intercorrelations, patient self-ratings and therapist ratings appear to covary to a high degree. Although measures of duration and course of illness have some relationship to improvement ratings, they seem also to tap different sources of variance.

Reliability. The reliability of outcome criteria has received attention; the duration and course measures are objective enough so that their reliability has been taken for granted. Miles, Barrabee, and Finesinger (1951) reported low interjudge but high test-retest intrajudge reliability of global judgments of improvement. Ten cases were rated by four judges on a six-point scale. There was complete agreement for only 20% of the judgments, though no disagreement was by more than two points. Test-retest figures showed 70% to 74% complete agreement between ratings taken 6 to 8 months apart. The ratings were based on structured interview material, and probably represent the lower bounds of interjudge agreement, if it is assumed that ratings made after a long period of observation of the patient would show

more stability than ratings made on the minimal information contained in a structured interview. These investigators felt that changes in psychiatric status over time cannot be discriminated any more finely than in terms of three gross classes: unchanged or worse, improved, and markedly improved. Levitt (1957) presented data suggesting that judged improvement rate tends to increase as a function of the number of points on the scale. The greatest discrepancy was due to studies using a two-point "improved-unimproved" scale, where the mean percentage improved was 51. Studies using three- to five-point scales had mean improvement rates of 73% to 76%.

A possible source of unreliability in judgments of improvement lies in the fact that they may confound the amount of change with the absolute level of terminal adjustment. Thus it seems likely that the reason initial severity of illness correlates with improvement (Conrad, 1954) is to some extent due to the fact that those who are high on a measure of adjustment initially will be high on adjustment terminally, though the change may be far from being as dramatic as for patients who are admitted in a state of confusion and disorientation, and discharged without these symptoms. Since each judge can combine amount of change and absolute level as he chooses, in most studies, a lowering of interjudge agreement is to be expected. This may be involved in the much higher interrater reliabilities reported by Morton (1955) than by Miles, Barrabee, and Finesinger (1951). Morton developed seven-point scales of absolute level of adjustment in 12 different areas. After training, the interrater reliability coefficients ranged from .79 to .91 when the ratings were based only on transcriptions of a terminal interview; and the reliability of the improvement score (the difference between ratings of an initial and terminal interview) ranged from .59 to .78.

Tests as criteria. A possible criterion of outcome is performance as measured by tests. The present review uncovered no studies which used changes in test scores as primary prognostic criteria but it remains a reasonable possibility. The primary requisite for this use of tests would be evidence that the tests covary with the changes in patients that go to make up the concept of improvement. A number of studies have been published which tackle this question, and in general they support the assumption of covariation.

Pascal and Zeaman (1951) found that the Bender-Gestalt, color-naming, noun-naming, and serial subtraction, from a larger battery of tests, correlated with the course of progress as judged clinically, for four patients getting electroconvulsive therapy.

Hybl and Stagner (1952) reported a significantly greater decrease in the amount of disruption of performance brought about by a frustration experience, for patients rated by their therapists as improved. The tasks were three psychomotor tests: the Ferguson Form Boards, Digit Symbol from the Wechsler-Bellevue, and the Minnesota Rate of Manipulation Test.

Vinson (1952) administered a mirror drawing test before and during electroshock therapy to 18 inpatients. Change in the mirror drawing score correlated .72 with change in orientation as evaluated by the clinical staff.

Several studies (Hozier, 1959; Wechsler, 1958) indicate that as psychotic patients improve there is a

decrease in variability of both the quality and the quantity of test performance.

The MMPI has been used in a number of studies of change: several studies (Carp, 1950; Feldman, 1951; Schofield, 1950, 1953) have reported that hospitalized patients treated with somatic therapies show an average drop on all of the MMPI scales of from 8 to 13 *T*-scale points. The acutely ill changed more than the chronically ill, and the affective disorders showed a greater change than the schizophrenics. Feldman (1951, 1952) found that improved patients' MMPI profiles dropped more than unimproved patients' profiles, and that the averaged profiles of these two groups showed greater differences after therapy than before. Work with predominantly psychoneurotic samples (Barron & Leary, 1955; Kaufman, 1950; Schofield, 1950) has indicated a larger drop on most scales for improved patients than for those rated unimproved. Changes taken without regard to sign (decreases as well as increases) were significantly greater in an individually treated group than in a group treated by group-therapy methods (Barron & Leary, 1955; Leary & Harvey, 1956).

Harris (1959) has summarized such MMPI studies to date as follows:

> scores on the MMPI show little change in normals and in untreated psychiatric patients over extended periods of time; somatic therapy, which is known to be effective at least in readying patients for discharge from the hospital, is accompanied by sizeable drops in test scores; patients in psychotherapy show smaller changes, perhaps not much larger than those produced by the passage of time alone; and the magnitude of change in test scores is related to clinical estimates of improvement (p. 519). (Quoted by permission of National Academy of Sciences–National Research Council)

Extraneous effects in test-retest comparisons need to be kept in mind, and Windle (1954) has reviewed these in reference to questionnaires. He presents evidence for a general tendency toward less deviant answers on retest, irrespective of external factors. This tendency is less, the greater the time period between test administrations. But even taking these artifactual sources of error into account, there appears to be evidence that a variety of test responses change in a manner consistent with therapist judgments of change in mental health.

Research in Prognosis

This section is organized around the three elements that seem most prominent in any treatment: the treatment itself, the person administering the treatment, and the patient who receives the treatment. Duration, course, or outcome of illness can potentially be affected by any one of these. The practical need to determine the prognosis of a patient implies that some selection is possible concerning the most appropriate treatment for that patient, or the most appropriate patient for a given treatment. Thus in the headings below we use the terms: treatment selection, therapist selection, and patient selection.

Treatment Selection

Ideally, the basic problem in prognosis is the assignment of patients to treatments in such a way as to maximize the total ratio of improved to unimproved patients. In decision theory terms, the prognostic judgment is a case of decision-making under conditions of certainty, which implies that the relationships between treatments and effects or outcomes are known. However, it has

not been demonstrated that different treatments have different effects. To quote an authority,

> One is reluctantly forced to admit that we simply do not possess the factual knowledge as of 1957 which permits us to say that we have any treatment procedure in psychiatry which promises a better outlook for a particular illness than does nature left to her own devices (Hastings, 1958, p. 1057). (Quoted by permission of the *American Journal of Psychiatry*)

Several attempts have been made to survey the literature on treatment effects, all of them hampered by the difficulties in comparing studies with different diagnostic groups, and different criteria for improvement. Eysenck (1952) selected 24 studies on the effect of psychotherapy with psychoneurotics, and concluded that these relatively homogeneous studies did not offer any evidence that improvement rate for those receiving psychotherapy was greater than for those getting only custodial care. Methodological weaknesses in his survey were pointed out by Rosenzweig (1954) and DeCharms, Levy, and Wertheimer (1954).

Levitt (1957) surveyed 30 articles evaluating psychotherapy with children. He compared the improvement rate on discharge and follow-up for treated cases with that reported for children accepted for therapy who never appeared for a first interview. The results were similar to those found by Eysenck, and did not demonstrate any facilitation of recovery due to psychotherapy.

Appel, Myers, and Scheflen (1953) summarized the results of studies which met a list of what they felt were minimal standards. They broke down the findings separately for schizophrenic, affective, and psychoneurotic disorders. Their survey indicated that none of the treatments studied—insulin coma, electroconvulsive shock, electronarcosis, lobotomy, or psychotherapy—gave recovery rates significantly greater than that reported for groups receiving only routine hospital care, in any of the three disorder categories. A more recent review by Staudt and Zubin (1957) covering the somatotherapies indicated that insulin and electroconvulsive shock temporarily increase the improvement rate, but after 3 years the increase has dissipated. This conclusion would seem to fly in the face of the fact that most of the studies reviewed by Staudt and Zubin reported significantly greater recovery for the treated group than for the control group at all periods of follow-up. However, the groups were equally different before treatment was begun; in most instances the control groups "seem to be highly selected and loaded with patients of apparently poor prognosis. Their improvement rates fall far short of the 'spontaneous improvement rates' " (Zubin, 1959, p. 344). This bias in selection of control groups is also likely to be operating in studies of psychotherapy unless matching procedures are possible, since there seems to be a feeling in many clinics that ethical considerations make it mandatory that patients who appear treatable be given treatment as quickly as possible.

Kramer and Greenhouse (1959) discuss a point which bears directly on the adequacy of studies in this area. They show the statistical implications of the common sense notion that the less dramatic the effect one is looking for, the larger the sample necessary to show that it is significant. Their tables indicate that if one is interested in identifying in the experimental group as slight an improvement as 5% over the control group (at the .05 level of significance) for base rate improvement of 40% (which is close to that found in

schizophrenia) it would take at least 569 cases in each group. For a base rate improvement of 70% (typical of the psychoneurotic) 472 cases per group would be needed to demonstrate a 5% increase under ideal conditions. These estimates further assume perfect reliability of the improvement criterion. Kramer and Greenhouse point out that very few states have a large enough population of mentally ill to do a study with a sample sufficient to detect slight but significant effects. Thus all the studies on the effect of treatment using small samples implicitly assume no interest in detecting anything less than extremely large differences. This is why it has been emphasized that treatment effects seem to be negligible relative to other variables in determining outcome; in view of the size of samples for research in this area it would not be fair to say that slight treatment effects may not exist.

How do patients regard psychotherapy? Stotsky (1956a) found that only 10% of a VA sample mentioned psychotherapy when asked to list any treatments which helped them. If asked directly whether they felt psychotherapy was the most important part of their treatment, over 50% said yes. These patients came predominantly from a lower socioeconomic class which, as will be discussed later, would bias the results in the direction of more negative answers.

Two final points can be made. It first should be said that clear-cut effects of psychotherapy seem to have been demonstrated using the patient's verbal behavior, rather than judgments of improvement, as the criterion measure (Rogers & Dymond, 1954; Rosenthal, 1955).

Secondly, it might be pointed out that the inconclusive state of affairs regarding the effects of treatment is not necessarily discouraging from the restricted point of view of the researcher. If treatment effects are currently less important than effects due to other sources of variance, then the researcher can ignore treatment differences in his samples and in the formulation of his hypotheses, thus considerably simplifying the research design.

Therapist Selection

A special aspect of treatment selection is the question of what kind of therapist does best with what kind of patient in psychotherapy. In the years surveyed in this review the pertinent articles in this area dealt with such therapist variables as sex, vocational interests, professional affiliation, and experience.

Irrespective of cause, are there differences between therapists as to treatment results? Imber, Frank, Nash, Stone, and Gliedman (1957) compared three therapists, each of whom worked with 18 patients. No significant differences were found between therapists, against a criterion of ratings of improvement in social effectiveness. Sullivan, Miller, and Smelser (1958) found neither sex, experience, nor profession (psychiatrist, psychologist, or social worker) to be related to either length of stay in therapy or to ratings of improvement. Hiler (1958a) reported significant differences in number of responses on the Rorschach between six groups of patients (14 per group), each group subsequently treated by a different therapist. He interpreted this as indicating that the therapists differed in their ability to keep unproductive patients in therapy. Stieper and Wiener (1959) found significant differences between therapists as to the length of time they kept patients in therapy. The differ-

ences seemed to be related to personality variables in the therapist, such as having high goals concerning very sick patients, and needing to feel appreciated. They took a negative view toward this minority of therapists who keep patients in therapy for long periods:

> It seems to us likely that psychotherapeutic practice today contains self-defeating concepts which may not only be hampering to the success of treatment, but potentially harmful to its clients (p. 241).

Betz and Whitehorn (1956) found differences in treatment between therapists who had a cumulatively high improvement rate with schizophrenics and therapists with a low improvement rate. The successful therapists were more active, emphasized utilization of assets, understood the meaning of the patient's behavior, and engendered more trust and confidence. They also differed from unsuccessful therapists in their scores on the Strong Vocational Interest Test.

Myers and Auld (1955) found that the experienced staff in an out-patient clinic had fewer patients quit against the therapist's wishes, and more patients who improved, than the residents in the same clinic. Katz and Solomon (1958) concluded that in their sample the less experienced therapists tended to lose more patients, but if the patient continued treatment, the improvement rate was as high as for the more experienced therapists. Strupp (1958) had 134 residents and psychiatrists respond to a sound film of an initial interview. He interpreted his data as showing two types of therapists. Type I was positive in his feelings toward the patient, optimistic about prognosis, and permissive and passive in therapy—and relatively inexperienced. Type II was more experienced, was negative toward the patient, pessimistic about prognosis, and active in therapy (giving orders and advice, and venting his irritations). Strupp quotes Kubie (1956) on reasons for this increasing pessimism: Kubie mentions his disappointment, saying it is one shared by other psychoanalysts, to find that with increasing experience he did not seem to have increasing success.

Several studies (Katz, Lorr, & Rubinstein, 1958; Sullivan et al., 1958) have reported that the more experienced the therapists, the larger the percentage of cases rated by him as improved; and the less severe the illness, the greater the likelihood of a patient's having an experienced therapist. Clearly, it is advisable to control for severity of illness in research on therapy. Differences in socioeconomic level also appear to interact with experience. Schaffer and Myers (1954) studied all cases accepted for treatment in an out-patient clinic during 1 year and found that

> the higher a patient's social class position . . . in the community, the greater were his chances of being accepted for psychotherapy, of being assigned to a relatively experienced therapist occupying a high status within the clinic, and of maintaining contact with the clinic (p. 88). (Quoted by permission of *Psychiatry*)

It is apparently also likely (Winder & Hersko, 1955) that the higher the social position, the higher the likelihood that the therapist will decide on analytic rather than supportive procedures.

Since the above studies did not control for these contaminating factors, it must be concluded that demonstration of between-therapist effects on outcome has not been conclusively obtained. This is not particularly surprising, in view of the fact that therapist selection is just a special case of treatment selection. Again, though, it can be said

that effects can probably be shown, against other than improvement criteria. For instance, Rosenthal (1955) found that the amount of benefit a client said he obtained from therapy correlated .68 with the degree of shift in moral values toward those held by the therapist, if the values had been talked about during psychotherapy. This change would appear to be related to those obtained in laboratory studies on verbal conditioning (Krasner, 1958).

Patient Selection: Outcome Criteria

We turn now to the question of the relationship of intra-individual variables to prognostic criteria. The studies will be grouped along two dimensions. They will be considered according to the kind of criterion used—outcome, duration, or course—and further broken down, where possible, in terms of the type of test used—projective, questionnaire, or performance (including cognitive tests).

Nontest indicators. Before turning to the research using tests as predictors of outcome, it is of interest to survey briefly what has been found using nontest variables. Huston and Pepernik (1958) reviewed prognostic variables in schizophrenia, and presented evidence that only these variables had been firmly established as going with favorable outcome: acute onset, short duration of illness prior to hospitalization, a precipitating stress, and the absence of flat or inappropriate affect. A series of studies under the direction of Pascal investigated the interrelationships of these variables within a sample of varied psychotics. It was found that acute onset (Swensen & Pascal, 1954b) and aggression directed toward oneself (Feldman, Pascal, & Swensen, 1954) related significantly to favorable outcome when other prognostic variables were controlled. However, precipitating stress (Cole, Swensen, & Pascal, 1954), affective expression (Bayard & Pascal, 1954), and duration of illness (Swensen & Pascal, 1954a) did not relate to outcome in their sample when the effect of other prognostic variables was held constant. The generality of their findings is not clear, since their method of balancing groups for control purposes led to their using only a small portion of the total sample, thus allowing for the possible introduction of unknown biases.

Eskey, Friedman, and Friedman (1957) could not find support for the notion that disorientation relates to duration of illness; however, they restricted their sample on the criterion variable by not using patients who were unimproved at discharge. Several studies (Eskey & Friedman, 1958; Phillips, 1953) indicate that intact cognitive processes and a mature premorbid social and sexual life go with favorable outcome. Zubin (1959) presents the results to date of an uncompleted survey of prognostic indicators for schizophrenia, which suggests that the variables defining reactive schizophrenia go with favorable prognosis, and those defining process schizophrenia go with unfavorable prognosis. He presents a valuable count of articles supporting or negating the postulated relationship for almost every if not every prognostic indicator that has been investigated. There have been several attempts to combine these variables into a scale. Thorne (1952) intuitively combined five into a quantified prognostic scale. More recently Lindemann, Fairweather, Stone, and Smith (1959) have developed a somewhat similar scale and cross-validated it against a criterion of duration of hospital stay. An eight-point scale (Schofield, Hatha-

way, Hastings, & Bell, 1954) developed to predict a follow-up criterion of adjustment in schizophrenia could not be cross-validated by Stone (1959). Becker and McFarland (1955) developed and cross-validated a 16-item scale against a criterion of improvement in a lobotomized sample.

The above studies have dealt with psychotics, or samples predominantly psychotic. Miles, Barrabee, and Finesinger (1951) reported that in a hospitalized psychoneurotic sample, age of onset, duration of illness prior to hospitalization, and a number of symptoms were unrelated to outcome. Patients with symptoms associated with autonomic discharge were most likely to remit. Rosenbaum, Friedlander, and Kaplan (1956), studying an outpatient sample, found improvement occurred in patients with good premorbid history whose environment offered many supports; and improvement was mainly in marital and work adjustment. Comparison of results on inpatient and outpatient samples suggests some reason for dealing separately with psychotics and psychoneurotics in prognosis research.

An important question is how well the clinician, using these nontest indices, can do in predicting outcome. Clow (1953) obtained a majority opinion of prognosis at the staff conference which was held 2 months after admission on each of 100 female schizophrenics. The prognoses were 73% correct in predicting a dichotomous improved-unimproved criterion obtained at discharge. More studies of this kind would be helpful in evaluating the practical usefulness of adding tests to current prognostic procedures.

Projective tests. Several Rorschach studies have used a configurational score, the Prognostic Rating Scale (PRS) (Klopfer, Kirkner, Wisham, & Baker, 1951). Kirkner, Wisham, and Giedt (1953) found a correlation of .67 between PRS and improvement ratings obtained by evaluating the terminal closure note, on a sample of 40 receiving psychotherapy. Mindess (1953) obtained a correlation of .66 (N of 70) between PRS and a diagnostic criterion running from normal through neurotic to psychotic, obtained 6 months after initiation of psychotherapy. Filmer-Bennett (1952, 1955) did not obtain significant results with either the PRS or global judgments based on the total Rorschach protocol. His criterion was a dichotomous improved-unimproved rating of the degree to which the patient was making a satisfactory social and vocational adjustment a year after discharge from the hospital. Rosalind D. Cartwright (1958) presented a review of several successful studies using the PRS, and described further positive results from her own study. The criterion was ratings of success of psychotherapy made by the counselor after termination of therapy. In an appended discussion of her paper Snyder argued that other tests might do as good a job with much less time needed for testing. Bloom (1956) added an interesting modification to his design. He divided his 46 subjects into two groups, an unproductive group (less than 11 Rorschach responses) and a productive group (11 or more responses). The PRS differentiated a dichotomous criterion of outcome of psychotherapy significantly in the productive group, but not in the unproductive. He further assessed 11 other scores, and found none which were either significant or nonsignificant for the total sample as a whole; all discriminated significantly in one or the other of his groups—four for the productive group, and seven

for the unproductive. His results suggest the operation of an interaction similar to the one Zubin and co-workers (see below) have reported between chronicity and outcome, and deserve further investigation.

Rogers and Hammond (1953) and Roberts (1954), both working with VA outpatients, tried a sign approach on the Rorschach with negative results. Dana (1954) hypothesized that Card IV, assumed to be most likely to pick up attitudes to authority, would give responses related to improvement in psychotherapy, if the authority relationship was crucial to outcome. The responses were placed in three categories—"adequate," "inadequate," "negative"—and there was a significant tendency for those with "adequate" response to improve, and those with "inadequate" responses to remain unimproved. Hammer (1953) felt that his review of the literature suggested that those patients whose Rorschach protocols look sicker than their H-T-P protocols have a good prognosis, while a poor prognosis is associated with giving more negative feelings on the H-T-P than on the Rorschach.

Ullman (1957) found two highly related measures—clinical judgments of TAT protocols and a social perceptions test—to be correlated significantly with two criteria of improvement: the Palo Alto Group Therapy Scale and hospital status after 6 months (hospitalized vs. discharged). S. Rosenberg (1954) developed and cross-validated eight prognostic signs based on the Wechsler-Bellevue, Sentence Completion, and on the Rorschach. Grauer (1953) found more Rorschach indices of anxiety in an improved group of schizophrenics than in an unimproved. Organic signs did not discriminate. The welter of signs which these studies find related to improvement shows no clear pattern. Obviously most of these positive findings with projective techniques should be further validated before they can be accepted as more than promising leads.

Questionnaires. Barron (1953b) reported lower pretherapy MMPI and Ethnocentrism scores for an improved outpatient group than for an unimproved group. The criterion was judgments of change in psychotherapy made by professionals who had not been involved in the treatment. At least some of these relationships were due to differences in IQ between the groups. Rosen (1954) was not able to verify Barron's finding with the E Scale. Barron developed a special ego strength scale from the MMPI (Barron, 1953a), which he successfully cross-validated against improvement criteria in three disparate samples. Wirt (1955, 1956) found the ego strength scale significantly discriminated an unimproved from a greatly improved group, the groups being extremes drawn from a hospitalized sample receiving psychotherapy. The scale did a better job of discrimination than experienced clinicians who based their judgments on the total MMPI profile.

Feldman (1951, 1952, 1958) explored the validity of the MMPI for the prediction of outcome after electroshock therapy. He found that items dealing with hostility and interpersonal relationships were predictive of outcome, while items dealing with symptomatology reflected the amount of improvement. Pumroy and Kogan (1958) were unable to cross-validate Feldman's prognostic scale in a small VA sample. Dana (1954) also obtained negative results with the MMPI, attempting to predict improvement after electroshock.

Performance tests. Stotsky (1956b) gave vocational aptitude and interest tests to a group of schizophrenics

most of whom had been in the hospital for a year or more. The aptitude tests predicted later work success, but the interest tests did not. Swensen and Pascal (1953) reported that the Pascal-Suttell Z score on the Bender-Gestalt test, was significantly lower for a group of inpatients judged to be improved on follow-up a year and a half later, than for those judged unimproved. Landis and Clausen (1955) found efficient performance on critical flicker fusion, reaction time, finger dexterity, auditory acuity threshold, and tapping speed was predictive of improvement in an inpatient sample receiving a variety of treatments. A variability score of palmar sweating (Ellsworth & Clark, 1957) predicted changes in a behavioral adjustment scale concurrent with the administration of tranquilizing drugs. Keehn (1955) took 12 measures from simple cognitive and psychomotor tests that had been shown to discriminate between normals and psychotics, and found only one score that predicted outcome in a group of inpatients receiving insulin coma therapy; he concluded that initial degree of psychoticism was not prognostic of outcome.

Vinson (1952) used a mirror drawing test to predict the prognosis made at discharge—a dichotomous "favorable-unfavorable" prognostic judgment made by the staff. His sample consisted of 18 hospitalized patients who received electroshock therapy. He tested before and during treatment, and the difference between these scores predicted the prognostic criterion at the .02 level of significance.

The most promising findings made in prognosis in the last 10 years have been reports coming out of the Columbia-Greystone project of two interaction effects. The first interaction dealt with the relation of chronicity to outcome. Windle and Hamwi (1953) reported that chronic patients who were discharged after treatment had poorer admission scores on a complex reaction time test than chronic patients who were not discharged. However, for acute patients, those whose illness was of short duration, the reverse was true, namely, poor admission scores were associated with poor outcome. Zubin, Windle, and Hamwi (1953) rechecked data on other tests, using chronic patients from the same study, and found four other tests which gave the same results. An independent validation was provided by Sonder (1955) using different tests. In all of these studies the results were most clear-cut for the chronic group, probably due to the fact that among the acute patients were some who were potentially or actually chronic.

The second interaction emerged from the study by Zubin, Windle, and Hamwi (1953) who found that the chronic patients who did well on conceptual tasks (intelligence, memory, personality tests) but poorly on perceptual tasks (learning and perception tests) had a poorer prognosis than chronic patients who showed conceptual confusion but perceptual clarity. Williams and Machi (1957), also working with the chronic sample from the Columbia-Greystone project, factor analyzed the test data, and found some support for this conceptual-perceptual differentiation. However, this finding is not yet as clearly supported by the evidence as the chronicity-outcome interaction. Zubin and Windle (1954) reviewed a number of independent prognostic studies, and reported that a consideration of the two interaction effects accounted for much of the conflicting findings. In the light of this work,

further attempts to investigate these interactions cannot help but be of value.

Patient Selection: Duration Criteria

Projective tests. Stotsky (1952), working only with schizophrenics, compared a group of patients who in a 2-year period had not left the hospital with a group which in the same period of time had been discharged and remained outside for at least 6 months. His hypothesis was that the prognosis would be best for patients with the best pretreatment emotional and intellectual integration. Of 19 Rorschach signs, 5 were significantly cross-validated in a second sample, Also, all of the 19 signs except *R* were found to be in the predicted direction in both samples.

Questionnaires. Grayson and Olinger (1957), in a VA inpatient sample, reported that those who were given early trial visits were able to give improved MMPIs when asked to respond in "the way a typical, well-adjusted person on the outside would do" to a greater extent than those still hospitalized after 3 months. Rapaport (1958) was not able to validate this finding, using a military sample, although the change on most of the scales was in the correct direction. Stieper and Wiener (1959) found a group of VA outpatients who were seen in psychotherapy for an average of 5.3 years had higher pretherapy scores on the MMPI scales, *Hs* and *Hy*, than a group who were discharged after 14 months.

A demographic study (Lindemann et al., 1959) found an index using marital status, diagnosis, degree of incapacity, legal competence, and alcohol intake as variables, was related to length of hospital stay. Ellsworth and Clayton (1959) found a rating scale of psychopathology filled out at admission did not correlate significantly with length of hospital stay, but a behavioral adjustment scale did correlate, patients with the best admission adjustment tending to remain in the hospital the shortest length of time.

Performance tests. Venables and Tizard (1956b) found "short-stay" schizophrenics performed better on a repetitive psychomotor task than did chronic schizophrenics. Reaction time differences (Venables & Tizard, 1956a) occurred on initial testing, but disappeared on retest.

Patient Selection: Course Criteria

Under criteria measuring the course of illness we have placed two broad questions: who will relapse, and who will terminate treatment.

Relapse. The broad question here is one of predicting who will get worse over time. It is of course the reverse of the question of who will improve. However, the prediction of improvement and its opposite may not necessarily be most effectively accomplished with the same test. It can not be assumed that the prediction of relapse or hospitalization can be made from the same tests which predict improvement. This is consistent with the assumption that change of mental status need not be a unitary concept.

Peterson (1954b) used the MMPI, Wechsler-Bellevue, Rorschach, and nontest data to predict who would require admission to the hospital from patients being seen on an outpatient basis in a VA mental hygiene clinic. Considering the base rates, the predictive power of the tests was slight, but the results suggested that the person who gets worse in therapy is single, has been previously hospitalized, is diagnosed psychotic, and has an MMPI profile strongly ele-

vated on the psychotic scales. Using a six-point scale based on signs of psychosis on the MMPI developed by Meehl, Peterson (1954a) was able to achieve 75% correct discrimination. Briggs (1958) was able to cross-validate this scale to a certain extent. He took patients who were already in the hospital when they received the MMPI. On follow-up he found the Peterson score differentiated those who were rehospitalized from those who were not only for patients originally diagnosed psychoneurosis or mixed psychoneurosis. This is consistent with Peterson's finding that in his study similar outpatient diagnoses were most often given to the cases which were later hospitalized.

Schofield and Briggs (1958) related several measures of improvement previous to initial discharge to rehospitalization, the median follow-up period being 5.8 years. Improvement in behavior ratings made by nurses was not related to rehospitalization, but a combination of ratings based on pre- and posttreatment MMPIs and psychiatric evaluations of improvement made at the time of discharge allowed 75% correct prediction for the 66% of cases on which the two ratings agreed. Since knowledge of the base rate alone would allow 66% correct prediction, this was only slightly better than chance.

Cowden, Deabler, and Feamster (1955), using a criterion of whether patient was rehospitalized within 90 days after discharge, reported judgments of change from admission to discharge on Sentence Completion and the H-T-P Test predicted the criterion. An "ego" score obtained from combining the Binet Vocabulary with Cards I, III, and VIII of the Rorschach predicted relapse within a 2-year period for a sample of discharge patients (Orr, Anderson, Martin, & Philpot, 1955), but did not predict discharge for a sample of nondeteriorated admissions. Working with a special group (outpatients considered interminable) Wiener (1959) studied return to psychotherapy over a 6-month period after initial psychotherapy was arbitrarily terminated. In his sample of 48, 37 returned for further therapy within this period. The MMPI did not discriminate returnees from nonreturnees. Months in treatment appeared to be a promising measure, with the returnees having a longer history of psychotherapy.

A study that fits under neither of our two course criteria is one by Rioch and Lubin (1959). They obtained lengthy follow-up data on 93 patients, sufficient to allow an assessment on an 11-point scale of how consistently the patient had moved upward or downward in his social adjustment over several years. Both the Wechsler-Bellevue IQ and a global rating based on the Rorschach correlated significantly with this criterion, mainly due to discrimination at the low end of the scale: all of the patients who deteriorated steadily had low scores on the predictors.

Termination of treatment. The criterion involved in the prediction of length of therapy is more objectively determined than improvement, but there are some difficulties in its determination nonetheless.

One question is how to measure length of therapy. Most studies have used the number of interviews as the measure. Number of weeks in treatment would appear to be an equivalent measure. However, Lorr, Katz, and Rubinstein (1958) found that the number of interviews correlated only .60 with number of weeks in treatment, and they argued that number of interviews is likely to be the less reliable of the two.

Another problem springs from the

research design used in most of the studies of termination. The total sample is usually divided into two groups, terminators and remainers, and test scores are related to this dichotomous criterion. The question becomes one of where to cut the distribution. Terminators have been defined as those remaining less than 4 sessions (Gliedman, Stone, Frank, Nash, & Imber, 1957), less than 10 sessions (Auld & Eron, 1953; Kotkov & Meadows, 1953), or less than 20 sessions (Gibby, Stotsky, Hiler, & Miller, 1954). Gibby et al. (1954) found that those terminating between 5–19 sessions resembled in their test responses those who terminated earlier rather than those continuing on for more than 19 sessions. Our previous discussion of the "failure zone" (Taylor, 1956) suggests that a variety of factors are operating in the first 20 weeks. When these factors have not been controlled, they can influence the findings in termination studies.

A further criticism has been made by Gundlach and Geller (1958) who suggest that termination rate and duration of illness are partly administrative artifacts, and partly a reflection of "the kind of personality problems that the staff are interested in, or skilled at, handling." This criticism can be taken as indirect support for the common practice of defining termination in terms of the distribution of the length of therapy measures, since in any given setting, the median or mean length takes some account of the effects of policy and staff interests.

Research on the prediction of termination by the use of *projective tests* shows a familiar, monotonous pattern: initial positive results with subsequent negative or indeterminate cross-validation. Kotkov and Meadow (1952, 1953) began with 12 formal scores, and validated one of these (*FC/CF*). They applied a formula based on three scores (*FC/CF*, *R*, *D%*) to another sample, and *D%* washed out. When these same signs are examined in an earlier study (Rogers, Knauss, & Hammond, 1951), none were significant, and only *R* was in the predicted direction. Auld and Eron (1953) tried a further validation of the Kotkov and Meadow formula, and obtained insignificant results. They found the Wechsler-Bellevue IQ accounted for the one Rorschach variable, *R*, which held up in their sample.

Starting anew, Gibby et al. (1954; Gibby, Stotsky, Miller, & Hiler, 1953) found 9 of 31 Rorschach signs promising. Taking the 9 to a second sample, 3 held up (*R*, *K*, *m*) and a predictive formula based on these variables was applied to a further independent sample, and afforded 68% correct prediction. However, knowledge of the base rate would have allowed 60% correct prediction, so the results were not strong enough to be of practical use. In their sample the Kotkov and Meadow formula did no better than chance, and IQ was not related to the criterion. Affleck and Mednick (1959) used an equation based on *R*, *M*, and *H* to predict who would remain for longer than three interviews. Their equation allowed 71% correct prediction in a validation sample. Their terminators were lower in IQ than the continuers (significant at .06 level). This is consistent with the findings of Auld and Eron (1953).

All of the above Rorschach studies except for Auld and Eron used equivalent VA males being seen on an outpatient basis, so in some respects sample homogeneity was better from study to study than is true of most validation research in this area. Of all the Rorschach signs only *R* seems to have maintained its promise in

these studies. More recent work (Gallagher, 1953, 1954; Taulbee, 1958) supports the conclusion that the number of Rorschach responses (*R*) relates to termination. However the Rorschach is probably an unnecessarily cumbersome way of measuring this variable; for instance, Gallagher (1954) found that the number of words used on the Mooney Problem Check List to describe the clients' problems was a better predictor than *R*.

Libo (1957) used a TAT-type test to predict the number of patients who would return the week after the test was administered. For 40 subjects he was able to make a significant prediction based on an "attraction score": the number of references in the stories to a desired move toward the therapist, or of anticipated satisfactions from therapy.

Three studies dealt with the prediction of termination in a tuberculosis hospital. Vernier, Whiting, and Meltzer (1955) were able to differentiate patients who left the hospital against medical advice from those who continued treatment to the end, using the Rorschach and H-T-P tests. The TAT did not discriminate. Moran, Fairweather, and Morton (1956), using a biographical inventory and an attitude questionnaire found that only prehospital life adjustment predicted who would leave the hospital prematurely, with those leaving having a long history of being unable to adjust to their life situations. Calden, Thurston, Stewart, and Vineberg (1955) developed and cross-validated a scale from the MMPI to predict premature discharge.

Taulbee (1958) developed a key based on the MMPI and the Rorschach to predict continuation of outpatient psychotherapy beyond the thirteenth interview. His results, not cross-validated, led him to conclude that those who continue in therapy are less defensive, and more persistent, dependent, anxious, and introspective than the terminators. Sullivan et al. (1958) reported no significant difference between MMPI scores of terminators, and continuers on a VA male sample. Of a number of variables only education and occupation related to the criterion. Conrad (1954) had therapists fill out a check list covering positive mental health, social conformity, and behavior pathology on VA outpatients with differing lengths of stay in psychotherapy. Continuers tended to look least disturbed initially, and to be at the median rather than at either extreme on social conformity.

Rubinstein and Lorr (1956) found differences between extreme groups (patients in psychotherapy for over 6 months vs. patients who had come less than six times and had terminated against the wishes of the therapist), on the authoritarian F Scale, and a vocabulary test. However, a later study (Lorr et al., 1958) which defined termination as having less than 7 weeks of psychotherapy, did not give significant results, though the scales were in the predicted direction. They combined a number of scales in a further attempt, and obtained a significant multiple correlation in a validation sample. However, the scales allowed no better prediction than interviewer's judgment.

A large recent project on termination was carried on at Johns Hopkins University (Frank, Gliedman, Imber, Nash, & Stone, 1957; Gliedman et al., 1957; Imber, Frank, Gliedman, Nash, & Stone, 1956; Imber, Nash, & Stone, 1955; Nash, Frank, Gliedman, Imber, & Stone, 1957). Their prognostic battery included an inventory and a Sway test. Those who stayed in

therapy more than three interviews were more suggestible on the Sway test, were more sociable, of higher socio-economic status, and more likely to see treatment as a means of maintaining status in their immediate social environment, than the terminators. When they compared group versus individual psychotherapy they found an interaction between treatment and termination: in group therapy, the terminators were more socially ineffective than the continuers, while the relationship was reversed for those getting individual therapy. This intriguing finding may have been related to an unequal distribution of social levels in the two groups—most of the lower class patients ended up in group psychotherapy, while most of the middle class patients were assigned to individual psychotherapy.

Hiler (1959) studied intial complaints, and concluded that continuers come to a clinic with typical psychoneurotic symptoms—obsessions, phobias, anxiety, depression, poor concentration—while early terminators are more likely to list purely organic symptoms, antisocial acts, or schizoid feelings. His continuers also obtained higher scores on the Wechsler-Bellevue with a subtest pattern characterized by Similarities being higher than Digit Span or Digit Symbol (Hiler, 1958b).

How much overlap is there between predictors of termination and improvement? Sullivan et al. (1958) investigated the relationship of MMPI scores and demographic variables to both improvement and termination criteria. Only occupational level was related significantly to both. Katz et al. (1958) found none of their predictors of length of stay correlated with therapist ratings of improvement. Frank et al. (1957) reported that a past history of social activity and a fluctuating course of illness was associated with continuation and improvement. A short duration of illness was associated with termination as well as improvement. Gallagher (1954) found the Taylor Manifest Anxiety Scale predicted continuation as well as improvement. In general the results suggest little overlap. This is somewhat unexpected, since as was mentioned earlier, there appears to be a positive relationship between criteria of duration of treatment and improvement. The most tenable assumption would seem to be that the variance shared by the two criteria is different from the variance shared by predictor and criterion. Possibly the correlation between criteria is due to rater bias.

Discussion

The previous sections of this paper have included the word "selection" in order to underline the fact that the practical need to predict to any of these criteria exists only when some sort of selection is necessary. For example, if the waiting list of an outpatient clinic is too long, selection of cases to receive treatment can be made on the basis of predicted probability of improving or terminating. If there is no need to deny treatment to anyone, knowledge of these prognostic probabilities is of no practical use. In most mental treatment centers today administrative procedures probably do not involve rejection of the patient as an alternative action, except in some outpatient clinics. Prognosis would be indispensable in the question of treatment selection, if differential effects of treatment were known; our survey has suggested that such effects have not yet been demonstrated. Thus it could be argued that prognosis is a sleeping giant at the present time, awaiting a future chance to be of service. Several other uses can be made of prognostic infor-

mation, of course. Knowledge of the variables which relate to changes in duration, course, or outcome of mental illness is of theoretical importance, an aid to understanding. A second promising use has been proposed by Feldman (1952) and Zubin (1959). They recommend that in nonprognostic research prognostic status be tried as a method of classifying patients into homogeneous categories, in place of diagnosis.

Is such a suggestion tantamount to substituting a measure of severity of illness for one of type of illness? The literature survey indicates a wide variety of tests have shown positive results, with no discernible common characteristic except that they measure adequacy of functioning, directly or indirectly. The fact that the same measures do not predict for all patients may be due to differences in the type and etiology of symptoms from patient to patient; but such differences do not vitiate the possibility that when prediction occurs it is largely because the dimension of severity of illness has been accurately assessed by the test. In any case, the effect of matching groups on prognostic variables would be to control for base rate differences in improvement, a procedure which is imperative for many kinds of evaluational studies, though rarely invoked in research on therapy.

As with all predictive questions, the primary problem in prognosis is the definition of the criterion. From the point of view of decision theory, the general notion of "outcome of illness" involves assigning utility values to specific outcomes; and since cost of achieving any given outcome may be a factor, an explication of the treatment strategies is also necessary. The low interjudge reliabilities which obtain in judgments of improvement indicate that utility of outcome may differ from judge to judge. A program for achieving a more objective ranking of treatments, outcomes, or treatment-outcome combinations seems called for. Cronbach and Gleser (1957) offer a possible framework for such a program, and most of the points they make, although dealing with personnel selection, can be easily generalized to prognosis.

A frequent misinterpretation of empirical research is that it is based on no theory. In the sense of a content theory—i.e., a theory stating relationships between tangibles or concepts related to tangibles—empirical research is usually weak, though in the selection of measures some sort of rough theory has to be involved. However, empirical research often is strongly tied to a mathematical model. In prognosis the guiding model has been the linear regression model. The studies have assumed that a measurable quality exists which is linearly related to outcome. The findings in respect to performance differences between acute and chronic patients (Burdock et al., 1958) suggest that this linear model probably will have to take account of interaction effects. If so, almost all studies to date are too simple in design. They involve a one-stage decision: look at one final score per person (the final score may of course be a combination of several subsidiary scores) and assign the patient to an outcome (criterion category) by whatever rule of operation is being applied to the score. The work of Zubin's group indicates that at least a two-stage decision process is needed: (*a*) a score is obtained to decide which of several operations will be applied to a second score, and (*b*) the second score is used to assign patients to the criterion category. Indeed there is no reason why tests should not be useful as a basis

for deciding what operational rule to apply to other data. The variables which appear to have the strongest relationship to outcome have been nontest variables: severity and duration of illness, acuteness of onset, degree of precipitating stress, etc. A possible direction of research might be to use tests to increase the validity of the nontest variables, either by trying to find tests which tap interactions, or which correlate with the error term in the psychiatric predictor. This latter approach has not been tried in prognosis, but it has been used with some success in personnel selection (Fulkerson, 1959; Ghiselli, 1956). A third suggested avenue of research would be to apply nonlinear or configurational models to prognostic data. The general point to be made is that prognosis research seems to require a different, more complex, mathematical model, and thus a more complex research design, than has been generally used so far. Specifically the one-stage design, where a predictor is correlated with an outcome measure, would appear to be inadequate in this field.

REFERENCES

AFFLECK, D. C., & MEDNICK, S. A. The use of the Rorschach test in the prediction of the abrupt terminator in individual psychotherapy. *J. consult. Psychol.*, 1959, **23**, 125–128.

APPEL, K. E., MYERS, J. M., & SCHEFLEN, A. E. Prognosis in psychiatry. *AMA Arch. Neurol. Psychiat.*, 1953, **70**, 459–468.

AULD, F., JR., & ERON, L. D. The use of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, **17**, 104–109.

AULD, F., JR., & MURRAY, E. J. Content-analysis studies of psychotherapy. *Psychol. Bull.*, 1955, **52**, 377–395.

BAILEY, M. A., WARSHAW, L., & EICHLER, R. M. A study of factors related to length of stay in psychotherapy. *J. clin. Psychol.*, 1959, **15**, 442–444.

BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333. (a)

BARRON, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241. (b)

BARRON, F., & LEARY, T. F. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 239–245.

BARRY, J. R. The relation of verbal reactions to adjustment level. *J. abnorm. soc. Psychol.*, 1950, **45**, 647–658.

BAYARD, JEAN, & PASCAL, G. R. Studies of prognostic criteria in the case records of hospitalized mental patients: Affective expression. *J. consult. Psychol.*, 1954, **18**, 122–126.

BECKER, W. C., & MCFARLAND, R. L. A lobotomy prognosis scale. *J. consult. Psychol.*, 1955, **19**, 157–162.

BETZ, BARBARA J., & WHITEHORN, J. C. The relationship of the therapist to the outcome of therapy in schizophrenia. *Psychiat. res. Rep.*, 1956, No. 5, 89–105.

BLOOM, B. L. Prognostic significance of the underproductive Rorschach. *J. proj. Tech.*, 1956, **20**, 366–371.

BRIGGS, P. F. Prediction of rehospitalization using the MMPI. *J. clin. Psychol.*, 1958, **14**, 83–84.

BURDOCK, E. I., SUTTON, S., & ZUBIN, J. Personality and psychopathology. *J. abnorm. soc. Psychol.*, 1958, **56**, 18–30.

BURDOCK, E. I., & ZUBIN, J. A rationale for the classification of experimental techniques in abnormal psychology. *J. gen. Psychol.*, 1956, **55**, 35–49.

CALDEN, G., THURSTON, J. R., STEWART, B. M., & VINEBERG, S. E. The use of the MMPI in predicting irregular discharge among tuberculosis patients. *J. clin. Psychol.*, 1955, **11**, 374–377.

CARP, A. MMPI performance and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 721–726.

CARTWRIGHT, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.

CARTWRIGHT, ROSALIND D. Predicting response to client-centered therapy with the Rorschach *PR* scale. *J. counsel. Psychol.*, 1958, **5**, 11–15.

CLOW, H. E. The use of a prognostic index of capacity for social adjustment in psychiatric disorders. In P. H. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis*. New York: Grune & Stratton, 1953. Pp. 89–106.

COLE, MARY E., SWENSEN, C. H., & PASCAL, G. R. Prognostic significance of precipitat-

ing stress in mental illness. *J. consult. Psychol.*, 1954, **18**, 171–175.

CONRAD, DOROTHY C. The duration of the therapeutic relationship and therapists' successive judgments of patients' mental health. *J. clin. Psychol.*, 1954, **10**, 229–233.

COWDEN, R. C., DEABLER, H. L., & FEAMSTER, J. H. The prognostic value of the Bender-Gestalt, H-T-P, TAT, and Sentence Completion Test. *J. clin. Psychol.*, 1955, **11**, 271–275.

CRANDALL, A., ZUBIN, J., METTLER, F. A., & LOGAN, N. D. The prognostic value of "mobility" during the first two years of hospitalization for mental disorder. *Psychiat. Quart.*, 1954 **28**, 185–210.

CRONBACH, L. J., & GLESER, G. *Psychological tests and personnel decisions.* Urbana: Univer. Illinois Press, 1957.

DANA, R. H. The effects of attitudes towards authority on psychotherapy. *J. clin. Psychol.*, 1954, **10**, 350–353.

DECHARMS, R., LEVY, J., & WERTHEIMER, M. A note on attempted evaluations of psychotherapy. *J. clin. Psychol.*, 1954, **10**, 233–235.

ELLSWORTH, R. B., & CLARK, L. D. Prediction of the response of chronic schizophrenics to drug therapy: A preliminary report on the relationship between palmar sweat and the behavioral effects of tranquilizing drugs. *J. clin. Psychol.*, 1957, **13**, 59–61.

ELLSWORTH, R. B., & CLAYTON, W. H. Measurement of improvement in "mental illness." *J. consult. Psychol.*, 1959, **23**, 15–20.

ENDS, E. J., & PAGE, C. W. Group psychotherapy and concomitant psychological change. *Psychol. Monogr.*, 1959, **73**(10, Whole No. 480).

ESKEY, A., FRIEDMAN, GLADYS M., & FRIEDMAN, I. Disorientation as a prognostic criterion. *J. consult. Psychol.*, 1957, **21**, 149–151.

ESKEY, A., & FRIEDMAN, I. The prognostic significance of certain behavioral variables. *J. consult. Psychol.*, 1958, **22**, 91–94.

EYSENCK, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.

EYSENCK, H. J., GRANGER, G. W., & BRENGELMANN, J. C. *Perceptual processes and mental illness.* London: Institute of Psychiatry, 1957.

FELDMAN, DOROTHY A., PASCAL, G. R., & SWENSEN, C. H. Direction of aggression as a prognostic variable in mental illness. *J. consult. Psychol.*, 1954, **18**, 167–170.

FELDMAN, M. J. A prognostic scale for shock therapy. *Psychol. Monogr.* 1951, **65**(10, Whole No. 327).

FELDMAN, M. J. The use of the MMPI profile for prognosis and evaluation of shock therapy. *J. consult. Psychol.*, 1952, **16**, 376–382.

FELDMAN, M. J. An evaluation scale for shock therapy. *J. clin. Psychol.*, 1958, **14**, 41–45.

FILMER-BENNETT, G. Prognostic indices in the Rorschach records of hospitalized patients. *J. abnorm. soc. Pscyhol.*, 1952, **47**, 502–506.

FILMER-BENNETT, G. The Rorschach as a means of predicting treatment outcome. *J. consult. Psychol.*, 1955, **19**, 331–334.

FLEISHMAN, E. A., & HEMPEL, W. E. A factor analysis of dexterity tests. *Personnel Psychol.*, 1954, **7**, 14–32.

FLEISHMAN, E. A., & HEMPEL, W. E. Factorial analysis of complex psychomotor performance and related skills. *J. appl. Psychol.*, 1956, **40**, 96–104.

FRANK, J. E., GLEIDMAN, L. H., IMBER, S. D., NASH, E. H., & STONE, A. R. Why patients leave psychotherapy. *AMA Arch. Neurol. Psychiat.*, 1957, **79**, 283–299.

FULKERSON, S. C. Individual differences in response validity. *J. clin. Psychol.*, 1959, **15**, 169–173.

GALLAGHER, J. J. The problem of escaping clients in nondirective counseling. In W. U. Snyder, *Group report of a program of research in psychotherapy.* State College: Pennsylvania State College, Psychotherapy Research Group, 1953. Pp. 21–38. (Mimeo)

GALLAGHER, J. J. Test indicators for therapy prognosis. *J. consult. Psychol.*, 1954, **18**, 409–413.

GHISELLI, E. E. Differentiation of individuals in terms of their predictability. *J. appl. Psychol.*, 1956, **40**, 374–377.

GIBBY, R. G., STOTSKY, B. A., HILER, H. W., & MILLER, D. R. Validation of Rorschach criteria for predicting duration of therapy. *J. consult. Psychol.*, 1954, **18**, 185–191.

GIBBY, R. G., STOTSKY, B. A., MILLER, D. R., & HILER, H. W. Prediction of duration of therapy from the Rorschach test. *J. consult. Psychol.*, 1953, **17**, 348–354.

GLASS, A. J., RYAN, F. J., LUBIN, A., REDDY, C. V. R., & TUCKER, A. C. Factors influencing psychiatrists in the prediction of military effectiveness. *Walter Reed Army Inst. Res. res. Rep.*, 1956, No. WRAIR-186–56.

GLEIDMAN, L. H., STONE, A. R., FRANK, D. D., NASH, E., JR., & IMBER, S. D. Incentives for treatment related to remaining or improving in psychotherapy. *Amer. J. Psychother.*, 1957, **11**, 589–598.

GLESER, GOLDINE, HADDOCK, J., STARR, P., & ULETT, G. A. Psychiatric screening of flying personnel: Inter-rater agreement on

the basis of psychiatric interviews. *USAF Sch. Aviat. Med. proj. Rep.* 1954, No. 10.

GORDON, M. H., LINDLEY, S. B., & MAY, R. B. A criterion measure of within-hospital change in psychiatric illness. *J. clin. Psychol.*, 1957, **13**, 145–147.

GRAUER, D. Prognosis in paranoid schizophrenia on the basis of the Rorschach. *J. consult. Psychol.*, 1953, **17**, 199–205.

GRAYSON, H. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, **21**, 73–77.

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, **53**, 267–293.

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist*, 1959, **14**, 469–479.

GUNDLACH, R. H., & GELLER, M. The problem of early termination: Is it really the terminee? *J. consult. Psychol.*, 1958, **22**, 410.

HAMMER, E. F. The role of the H-T-P in the prognostic battery. *J. clin. Psychol.*, 1953, **9**, 371–374.

HARRIS, R. E. The prediction and measurement of drug-induced psychological change. In J. O. Cole & R. W. Gerard (Eds.), *Psychopharmacology: Problems in evaluation.* (NAS-NRC Publ. No. 583) Washington, D. C.: National Academy of Sciences–National Research Council, 1959. Pp. 514–528.

HASTINGS, D. W. Follow-up results in psychiatric illness. *Amer. J. Psychiat.*, 1958, **114**, 1057–1066.

HEMPEL, W. E., & FLEISHMAN, E. A. A factor analysis of physical proficiency and manipulative skills. *J. appl. Psychol.*, 1955, **39**, 12–16.

HILER, E. W. An analysis of patient-therapist compatibility. *J. consult. Psychol.*, 1958, **22**, 341–347. (a)

HILER, E. W. Wechsler-Bellevue intelligence as a predictor of continuation in psychotherapy. *J. clin. Psychol.*, 1958, **14**, 192–194. (b)

HILER, E. W. Initial complaints as predictors of continuation in psychotherapy. *J. clin. Psychol.*, 1959, **15**, 244–245.

HOZIER, Ann. On the breakdown of the sense of reality: A study of spatial perception in schizophrenia. *J. consult. Psychol.*, 1959, **23**, 185–194.

HUSTON, P. H., & PEPERNIK, M. C. Prognosis in schizophrenia. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome.* New York: Logos, 1958. Pp. 531–546.

HYBL, A. R., & STAGNER, R. Frustration tolerance in relation to diagnosis and therapy. *J. consult. Psychol.*, 1952, **16**, 163–170.

IMBER, S. D., FRANK, J. D., GLIEDMAN, L. H., NASH, E. H., & STONE, A. R. Suggestibility, social class and the acceptance of psychotherapy. *J. clin. Psychol.*, 1956, **12**, 341–344.

IMBER, S. D., FRANK, J. D., NASH, E. H., STONE, A. R., & GLIEDMAN, L. H. Improvement and amount of therapeutic contact: An alternative to the use of no-treatment controls in psychotherapy. *J. consult. Psychol.*, 1957, **21**, 309–315.

IMBER, S. D., NASH, E. H., & STONE, A. R. Social class and duration of psychotherapy. *J. clin. Psychol.*, 1955, **11**, 281–284.

JAHODA, MARIE. *Current concepts of positive mental health.* New York: Basic Books, 1958.

KALIS, BETTY L., & BENNETT, LILLIAN F. The assessment of communication: The relation of clinical improvement to measured changes in communicative behavior. *J. consult. Psychol.*, 1957, **21**, 10–14.

KATZ, J., & SOLOMON, REBECCA Z. The patient and his experiences in an outpatient clinic. *AMA Arch. Neurol. Psychiat.*, 1958, **80**, 86–92.

KATZ, M. M., LORR, M., & RUBINSTEIN, E. A. Remainder patient attributes and their relation to subsequent improvement in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 411–413.

KAUFMAN, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric therapy. *J. consult. Psychol.*, 1950, **14**, 458–464.

KEEHN, J. D. An investigation into the value of "objective test psychoticism" in predicting response to insulin coma therapy. *J. ment. Sci.*, 1955, **101**, 871–877.

KELMAN, H. C., & PARLOFF, M. B. Interrelations among three criteria of improvement in group therapy: Comfort, effectiveness, and self-awareness. *J. abnorm. soc. Psychol.*, 1957, **54**, 281–288.

KING, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954.

KIRKNER, F. J., WISHAM, W. W., & GIEDT, F. H. A report on the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 465–470.

KLOPFER, B., KIRKNER, F. J., WISHAM, W., & BAKER, G. Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1951, **15**, 425–428.

KNIGHT, R. P. Evaluation of the results of psychoanalytic therapy. *Amer. J. Psychiat.*, 1941, **98**, 434–446.

KOTKOV, B., & MEADOW, A. Rorschach criteria for continuing group psychotherapy. *Int. J. group Psychother.*, 1952, **2**, 324–333.

KOTKOV, B., & MEADOW, A. Rorschach cri-

teria for predicting continuation in individual psychotherapy. *J. consult. Psychol.*, 1953, **17**, 16–20.

KRAMER, M., & GREENHOUSE, S. W. Determination of sample size and selection of cases. In J. O. Cole & R. W. Gerard (Eds.), *Psychopharmacology: Problems in evaluation.* (NAS-NAC Publ. No. 583) Washington, D. C.: National Academy of Sciences–National Research Council, 1959. Pp. 356–371.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 145–170.

KUBIE, L. S. Some unsolved problems of psychoanalytic psychotherapy. In F. Fromm-Reichman & J. L. Moreno (Eds.), *Progress in psychotherapy 1956.* New York: Grune & Stratton, 1956. Pp. 87–102.

LANDIS, C., & CLAUSEN, J. Changes in sensory and motor performances induced by active psychiatric treatment. *J. Psychol.*, 1955, **40**, 275–305.

LANGFELDT, G. The prognosis in schizophrenia. *Acta psychiat. neurol. Scand., Kbh.*, 1956, Suppl. 110, 1–66.

LEARY, T., & HARVEY, J. S. A methodology for measuring personality changes in psychotherapy. *J. clin. Psychol.*, 1956, **12**, 123–132.

LEVITT, E. E. The results of psychotherapy with children: An evaluation. *J. consult. Psychol.*, 1957, **21**, 189–196.

LIBO, L. M. The projective expression of patient-therapist attraction. *J. clin. Psychol.*, 1957, **13**, 33–36.

LINDEMANN, J. H., FAIRWEATHER, G. W., STONE, G. B., & SMITH, R. S. The use of demographic characteristics in predicting length of neuropsychiatric hospital stay. *J. consult. Psychol.*, 1959, **23**, 85–89.

LORR, M., KATZ, M. M., & RUBINSTEIN, E. A. The prediction of length of stay in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 321–327.

LUCE, R. D., & RAIFFA, H. *Games and decisions.* New York: Wiley, 1957.

MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.* 1955, **52**, 194–216.

MILES, H. H. W., BARRABEE, E. L., & FINESINGER, J. E. Evaluation of psychotherapy with a follow-up study of 62 cases of anxiety neurosis. *Psychosom. Med.*, 1951, **13**, 83–105.

MINDESS, H. Predicting patients' responses to psychotherapy: A preliminary study designed to investigate the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 327–334.

MORAN, L. J., FAIRWEATHER, G. W., & MORTON, R. B. Some determinants of successful and unsuccessful adaptation to hospital treatment of tuberculosis. *J. consult. Psychol.*, 1956, **20**, 125–131.

MORSE, P. W. A proposed technique for the evaluation of psychotherapy. *Amer. J. Orthopsychiat.*, 1953, **23**, 716–731.

MORTON, R. B. An experiment in brief psychotherapy. *Psychol. Monogr.*, 1955, **67**(1, Whole No. 386).

MYERS, J. K., & AULD, F. Some variables related to outcome of psychotherapy. *J. clin. Psychol.*, 1955, **11**, 51–54.

NASH, E. H., JR., FRANK, J. D., GLIEDMAN, L. H., IMBER, S. D., & STONE, A. R. Some factors related to patients remaining in group psychotherapy. *Int. J. group Psychother.*, 1957, **7**, 264–274.

ORR, W. F., ANDERSON, R. B., MARTIN, M. P., & PHILPOT, D. F. Factors influencing discharge of female patients from a state mental hospital. *Amer. J. Psychiat.*, 1955, **111**, 576–582.

PARLOFF, M. B., KELMAN, H. C., & FRANK, J. D. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *Amer. J. Psychiat.*, 1954, **111**, 343–351.

PASCAL, G. R., SWENSEN, C. H., FELDMAN, D. A., COLE, M. E., & BAYARD, J. Prognostic criteria in the case histories of hospitalized mental patients. *J. consult. Psychol.*, 1953, **17**, 163–171.

PASCAL, G. R., & ZAX, M. Psychotherapeutics: Success or failure. *J. consult. Psychol.*. 1956, **20**, 325–331.

PASCAL, G. R., & ZEAMAN, JEAN B. Measurement of some effects of electroconvulsive therapy on the individual patient. *J. abnorm. soc. Psychol.* 1951, **45**, 104–115.

PETERSON, D. R. The diagnosis of subclinical schizophrenia. *J. consult. Psychol.*, 1954, **18**, 198–200. (a)

PETERSON, D. R. Predicting hospitalization of psychiatric outpatients. *J. abnorm. soc. Psychol.*, 1954, **49**, 260–265 (b)

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

PUMROY, D. K., & KOGAN, W. S. A validation of measures that predict the efficacy of shock therapy. *J. clin. Psychol.*, 1958, **14**, 46–47.

RABIN, A. I., & KING, G. F. Psychological studies. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome.* New York: Logos 1958. Pp. 216–278.

RAPAPORT, G. M. "Ideal self" instructions, MMPI profile changes, and the prediction,

of clinical improvement. *J. consult. Psychol.*, 1958, **22**, 459–463.

RENNIE, T. A. C. Prognosis in the psychoneuroses: Benign and malignant developments. In P. H. Hoch & J. Zubin (Eds.), *Current problems in psychiatric diagnosis.* New York: Grune & Stratton, 1953. Pp. 67–79.

REZNIKOFF, M., & TOOMEY, L. C. *Evaluation of changes associated with psychiatric treatment.* Springfield: Charles C Thomas, 1959.

RIOCH, MARGARET J., & LUBIN, A. Prognosis of social adjustment for mental hospital patients under psychotherapy. *J. consult. Psychol.*, 1959, **23**, 313–318.

ROBERTS, L. K. The failure of some Rorschach indices to predict the outcome of psychotherapy. *J. consult. Psychol.*, 1954, **18**, 96–98.

ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

ROGERS, L. S., & HAMMOND, K. R. Prediction of the results of therapy by means of the Rorschach test. *J. consult. Psychol.*, 1953, **17**, 8–15.

ROGERS, L. S., KNAUSS, JOANNE, & HAMMOND, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368–371.

ROSEN, E. Ethnocentric attitude changes and rated improvement in hospitalized psychiatric patients. *J. clin. Psychol.*, 1954, **10**, 345–350.

ROSENBAUM, M., FRIEDLANDER, JANE, & KAPLAN, S. M. Evaluation of results of psychotherapy. *Psychosom. Med.*, 1956, **18**, 113–132.

ROSENBERG, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. clin. Psychol.*, 1954, **10**, 341–345.

ROSENTHAL, D. Changes in some moral values following psychotherapy. *J. consult. Psychol.*, 1955, **19**, 431–436.

ROSENZWEIG, S. A transvaluation of psychotherapy: A reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, **49**, 298–304.

RUBINSTEIN, E. A., & LORR, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–349.

SCHAFFER, L., & MYERS, J. K. Psychotherapy and social stratification: An empirical study of practice in a psychiatric out-patient clinic. *Psychiatry*, 1954, **17**, 83–93.

SCHOFIELD, W. Changes in response to the Minnesota Multiphasic Personality Inventory following certain therapies. *Psychol. Monogr.*, 1950, **64**(5, Whole No. 311).

SCHOFIELD, W. A further study of the effects of therapies on MMPI responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 67–77.

SCHOFIELD, W., & BRIGGS, P. F. Criteria of therapeutic response in hospitalized psychiatric patients. *J. clin. Psychol.*, 1958, **14**, 227–232.

SCHOFIELD, W., HATHAWAY, S. R., HASTINGS, D. W., & BELL, DOROTHY M. Prognostic factors in schizophrenia. *J. consult. Psychol.* 1954, **18**, 155–166.

SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 29–45.

SEASHORE, R. A., BUXTON, C. E., & MCCOLLUM, I. N. Multiple factorial analysis of fine motor skills. *Amer. J. Psychol.*, 1940, **53**, 251–259.

SEEMAN, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954. Pp. 99–108.

SNYDER, W. U. *Group report of a program of research in psychotherapy.* State College: Pennsylvania State College, Psychotherapy Research Group, 1953. (Mimeo)

SONDER, SYLVIA L. Perceptual tests and acute and chronic status as predictors of improvement in psychotic patients. *J. consult. Psychol.*, 1955, **19**, 387–392.

STANDAL, S. W., & VAN DER VEEN, F. Length of therapy in relation to counselor estimates of personal integration and other case variables. *J. consult. Psychol.*, 1957, **21**, 1–9.

STAUDT, VIRGINIA, M., & ZUBIN, J. A biometric evaluation of the somatotherapies in schizophrenia. *Psychol. Bull.*, 1957, **54**, 171–196.

STIEPER, D. R., & WIENER, D. N. The problem of interminability in outpatient psychotherapy. *J. consult. Psychol.*, 1959, **23**, 237–242.

STILSON, D. W., MASON, D. J., GYNTHER, M. D., & GERTZ, B. An evaluation of the comparability and reliabilities of two behavior rating scales for mental patients. *J. consult. Psychol.*, 1958, **22**, 213–216.

STONE, BETH. Prognostic factors in schizophrenia. *J. consult. Psychol.*, 1959, **23**, 279.

STORROW, H. A. The measurement of change in psychotherapy. *Scientific Papers of the 115th Annual Meeting of the American Psychiatric Association*, 1959, 41–42. (Abstract)

STORROW, H. A. The measurement of outcome in psychotherapy. *AMA Arch. gen. Psychiat.*, 1960, **2**, 142–146.

STOTSKY, B. A. A comparison of remitting and non-remitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 489–496.

STOTSKY, B. A. How important is psychotherapy to the hospitalized psychiatric patient. *J. clin. Psychol.*, 1956, **12**, 32–36. (a)

STOTSKY, B. A. Vocational tests as measures of performance of schizophrenics in two rehabilitation activities. *J. clin. Psychol.*, 1956, **12**, 236–242. (b)

STRUPP, H. H. The psychotherapists' contribution to the treatment process. *Behav. Sci.*, 1958, **3**, 34–67.

SULLIVAN, P. L., MILLER, CHRISTINE, & SMELSER, W. Factors in length of stay and progress in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 1–9.

SWENSON, C. H., & PASCAL, G. R. A note on the Bender-Gestalt test as a prognostic indicator in mental illness. *J. clin. Psychol.*, 1953, **9**, 398.

SWENSEN, C. H., JR., & PASCAL, G. R. Duration of illness as a prognostic indicator in mental illness. *J. consult. Psychol.*, 1954, **18**, 363–365. (a)

SWENSEN, C. H., JR., & PASCAL, G. R. Prognostic significance of type of onset of mental illness. *J. consult. Psychol.*, 1954, **18**, 127–130. (b)

TAULBEE, E. S. Relationship between certain personality variables and continuation in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 83–89.

TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.* 1955, **19**, 205–209.

TAYLOR, J. W. Relationship of success and length in psychotherapy. *J. consult. Psychol.*, 1956, **20**, 332.

THORNE, F. C. The prognostic index. *J. clin. Psychol.*, 1952, **8**, 42–45.

THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.

ULLMAN, L. P. Selection of neuropsychiatric patients for group psychotherapy. *J. consult. Psychol.*, 1957, **21**, 277–280.

VENABLES, P. H., & TIZARD, J. Paradoxical effects in the reaction time of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 220–224. (a)

VENABLES, P. H., & TIZARD, J. Performance of functional psychotics on a repetitive task. *J. abnorm. soc. Psychol.*, 1956, **53**, 23–26. (b)

VERNIER, C. M., WHITING, J. F., & MELTZER, M. L. Differential prediction of a specific behavior from three projective techniques. *J. consult. Psychol.*, 1955, **19**, 175–182.

VINSON, D. B., JR. Response to electroshock therapy as evaluated by mirror drawing. *J. clin. exp. Psychopath.*, 1952, **13**, 201–210.

VOSBURG, R. Some remarks on psychotherapy as reflected in hospital charts. *Psychiat. Commun.*, 1958, **1**, 151–160.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins, 1958.

WIENER, D. N. The effect of arbitrary termination on return to psychotherapy. *J. clin. Psychol.*, 1959, **15**, 335–338.

WILLIAMS, R. J., & MACHI, V. S. An analysis of interperson correlations among thirty psychotics. *J. abnorm. soc. Psychol.*, 1957, **55**, 50–57.

WINDER, A. E., & HERSKO, M. The effect of social class on the length and type of psychotherapy in a Veterans' Administration mental hygiene clinic. *J. clin. Psychol.*, 1955, **11**, 77–79.

WINDER, C. L. Psychotherapy. *Annu. Rev. Psychol.*, 1957, **8**, 309–330.

WINDLE, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451–482.

WINDLE, C. Test-retest effect on personality questionnaires. *Educ. psychol. Measmt.*, 1954, **14**, 617–633.

WINDLE, C., & HAMWI, VIOLET. An exploratory study of the prognostic value of the Complex Reaction Time Test in early and chronic psychotics. *J. clin. Psychol.*, 1953, **9**, 156–161.

WIRT, R. D. Further validation of the ego-strength scale. *J. consult. Psychol.*, 1955, **19**, 444.

WIRT, R. D. Actuarial prediction. *J. consult. Psychol.*, 1956, **20**, 123–124.

ZUBIN, J. A biometric model for psychopathology. In R. A. Patton (Ed.), *Current trends in the description and analysis of behavior.* Pittsburgh: Univer. Pittsburgh Press, 1958. Pp. 22–47.

ZUBIN, J. Role of prognostic indicators in the evaluation of therapy. In J. O. Cole & R. W. Gerard (Eds.), *Psychopharmacology: Problems in evaluation.* (NAS-NRC Publ. No. 583) Washington, D. C.: National Academy Sciences–National Research Council, 1959. Pp. 343–355.

ZUBIN, J., & WINDLE, C. Psychological prognosis of outcome in the mental disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 272–281.

ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.* 1953, **21**, 342–355.

(Received April 25, 1960)

Psychological Bulletin
1961, Vol. 58, No. 3, 205–217

COMPLEX SOUNDS AND CRITICAL BANDS[1]

BERTRAM SCHARF
Northeastern University

Studies of the responses of human observers to bands of noise and other complex sounds have led to the measure of what appears to be a basic unit of hearing, the critical band. When the frequency spectrum of a stimulating sound is narrower than the critical band, the ear reacts one way; when the spectrum is wider, it reacts another way. For example, experiments show that at values less than the critical bandwidth, both loudness and absolute threshold are independent of bandwidth; only when the critical bandwidth is exceeded do the loudness and the absolute threshold increase with the width (Gässler, 1954; Zwicker & Feldtkeller, 1955; Zwicker, Flottorp, & Stevens, 1957).

The critical band has also been measured in experiments on auditory discriminations that seem to depend upon phase (Zwicker, 1952) and in experiments on the masking of a narrow band of noise by two tones (Zwicker, 1954). In all four types of experiment—loudness, threshold, sensitivity to phase, and two-tone masking—the value of the critical band is the same function of its center frequency. The values of the critical band, as a function of the frequency at the center of the band, are given by the top curve in Figure 1. The ordinate gives the width (ΔF), in cycles per second, of the critical band; the abscissa gives the center frequency. As the frequency at the center of a complex sound increases, the critical band that is measured around the center frequency becomes wider.

Not only does the critical band have the same values when measured for several kinds of auditory response, it is also independent of such stimulus parameters as the number of components in the complex (Scharf, 1959b) and the sound pressure level (Feldtkeller, 1955; Feldtkeller & Zwicker, 1956).

Prior to the experimental measures of the critical band, Fletcher (1940) had hypothesized the existence of a critical band for masking. He suggested that when a white noise just masks a tone, only a relatively narrow band of frequencies surrounding the tone does the masking, energy outside the band contributing little or nothing. Although attempts to test this hypothesis remain inconclusive, investigators (Bilger & Hirsh, 1956; Hawkins & Stevens, 1950) have been able to calculate values for the width of these hypothetical masking bands by assuming that the masking band and the just-masked tone have the same intensity. The calculated values, which are labeled "critical ratios" in Figure 1, are smaller for the masking band than for the critical band as measured in the experiments cited above. As we shall see, this discrepancy is more apparent than real.

[1] This paper was completed while the author held a research grant, B2223, from the National Institute of Health, United States Public Health Service. The author wishes to thank S. S. Stevens for detailed criticism and advice in the final preparation of the manuscript. A. B. Warren and H. S. Zamansky are also thanked for reading and making suggestions about the final draft.

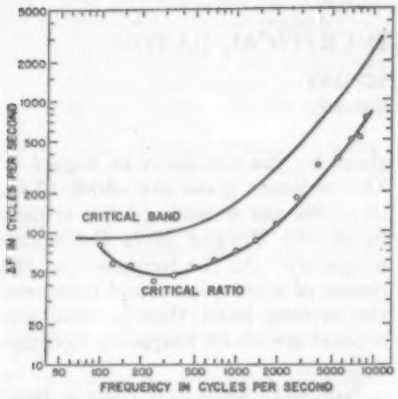

FIG. 1. The width, ΔF, of the critical band and of the critical ratio as a function of the frequency at the center of the band. (The ordinate gives the width, in cycles per second, of the critical band—and of the critical ratio—for the center frequencies shown on the abscissa. The top curve gives the values for the critical band which are based upon direct measurements in four types of experiment; the bottom curve gives the values for the critical ratio which are calculated from measurements of the masked threshold for pure tones in white noise. The points on the bottom curve are from Hawkins and Stevens—1950. This figure is adapted from an article by Zwicker, Flottorp, and Stevens—1957, p. 556—which contains also a table of critical-band values.) (Adapted with permission of the *Journal of the Acoustical Society of America*)

EXPERIMENTAL MEASURES OF THE CRITICAL BAND

Four types of experiment in which critical bands have been measured are reviewed: absolute threshold of complex sounds, masking of a band of noise by two tones, sensitivity to phase differences, and loudness.

Threshold of Complex Sounds

When two tones, whose frequencies are not too far apart, are presented simultaneously, a subject may report hearing a sound even though either tone by itself is below threshold. Gässler (1954) made careful measures of this phenomenon, using many tones and systematically varying the difference in frequency, ΔF, between the lowest and highest components of the complex sounds.[1] He varied the ΔF by varying the number of equally intense tones, which were spaced at intervals of 20 cps. The number of tones was increased from 1 to 40 or until ΔF was equal to 780 cps. Each time a tone was added, the threshold for the whole complex was measured by a "tracking" method (Stevens, 1958). It was necessary, of course, that all the tones in the complex have the same threshold when heard singly, for otherwise it would have been impossible to determine the precise cause of a change in the threshold for a complex whose ΔF had been increased by the addition of a tone. Thus measurements were restricted to portions of the frequency spectrum over which a subject's threshold curve was flat. In order to study other portions of the spectrum, the multitone complexes were presented against a background of white noise that had been tailored to raise the threshold for tones at all the audible frequencies to the same level, thus artificially flattening a subject's threshold curve.

Whether the background was quiet, or consisted of a noise at 0 db. SPL, at 20 db., or at 40 db., the same effect was noted: as soon as ΔF exceeded a particular value whose size depended upon the frequency at the center of the complex, the threshold for the multitone complex began to increase. Similar data were reported when bands of white noise were substituted for the multitone

[1] Two or more tones constitute a complex sound, i.e., a sound with energy at more than one frequency in contrast to a single or pure tone with most of its energy concentrated at a single frequency.

complexes. The results indicate that the *total* energy necessary for a sound to be heard remains constant so long as the energy is contained within a limiting bandwidth. Although differences between the two observers in these experiments were sometimes of the order of 40%, the average size of the limiting bandwidths for both multitone complexes and bands of noise is approximated by the critical-band curve of Figure 1.[3]

Two-Tone Masking

The masking of a narrow-band noise by two tones provided a second measure of the critical band. Using a tracking method, Zwicker (1954) measured the threshold of a narrow-band noise in the presence of two tones, one on either side of the noise. Increasing the difference in frequency, ΔF, between the two tones left the masked threshold for the noise unchanged until a critical ΔF was reached, whereupon the threshold fell sharply and, in general, continued to fall as ΔF was increased further. The two subjects who served in this experiment showed the same drop in threshold at approximately the same ΔF for a given center frequency regardless of the SPL of the masking tones. The critical-band curve of Figure 1 gives the approximate values of ΔF at which the masking effect of two tones is sharply reduced.

[3] Gässler (1954) measured a critical band of 165 cps at 1000 cps. Garner (1947) had written earlier that "the best estimate . . . is that a band of frequencies no wider than 175 cps around 1000 cps is necessary if temporal integration of acoustic energy is to be perfect" (p. 813). His estimate was based upon measurements of the threshold changes for a wide-band noise, an unfiltered 1000-cycle tone, and a filtered 1000-cycle tone as a function of bandwidth which was varied by varying the duration of the signal.

Sensitivity to Phase

The critical band is also relevant to phase sensitivity, measured by a comparison between the ear's ability to detect amplitude modulation (AM) and its ability to detect frequency modulation (FM). This procedure requires some explanation.

When the *amplitude* of a tone is modulated—i.e., alternately increased and decreased—a three-tone complex is produced with the original tone (the "carrier") at the center of the complex and a tone on either side (side bands). When the *frequency* of a tone is modulated over a narrow range, a three-tone complex is also produced.[4] The only important difference between the three-tone complex that is produced under AM and the complex that is produced under FM concerns the phase relations among the components. Consequently, any difference in the ear's sensitivity to AM and FM would presumably depend upon these phase relations.

Zwicker (1952) found, indeed, that in order for a subject to just hear a difference between a modulated and a pure, unmodulated tone, a smaller amount of AM is required than FM. The ear is more sensitive to AM than to FM, however, only at low rates of modulation. As the rate of modulation is increased, the difference in sensitivity to AM and FM gradually disappears. How do these results pertain to the critical band? The rate at which a tone is modulated determines the frequency separation, ΔF, between the side bands of the three-tone complex produced under the modulation. It turns out that the rate of modulation at which AM and

[4] For a lucid discussion of the intricacies of modulation, consult Stevens and Davis (1938, pp. 225–231).

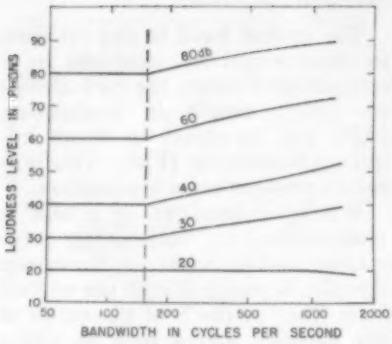

FIG. 2. The loudness level of a band of noise centered at 1000 cps measured as a function of the width of the band. (The parameter is the effective SPL of the noise. The dashed line shows that the bandwidth at which loudness begins to increase is the same at all the levels tested. This figure is adapted from the book, *Das Ohr als Nachrichtenempfänger*, by Feldtkeller and Zwicker—1956, p. 82.) (Adapted with permission of S. Hirzel Verlag)

FM become equally difficult to detect corresponds to values of ΔF that are essentially the same as the critical-band values given in Figure 1. Zwicker's investigation showed, moreover, that the critical band determined by phase sensitivity is independent of the SPL of the modulated tone and varies only as a function of the frequency of the "carrier" which lies, of course, at the center of the band.

Since the complexes produced under AM and those produced under FM differ primarily with respect to phase relations, the ear may be able to detect AM more easily than FM at low rates of modulation because it is more sensitive to the kind of phase relations that occur under AM. The ear seems to be sensitive to the phase relations, however, only when the ΔF of the complex is less than a critical band. When ΔF is greater than a critical band, there is no difference in sensitivity to AM and FM, implying that, beyond the critical band, the phase relations within the complex no longer serve as a significant cue in the detection of modulation.

Loudness of Complex Sounds

The critical band has been measured most thoroughly in studies of the loudness of complex sounds as a function of bandwidth. Zwicker and Feldtkeller (1955) demonstrated that the loudness of a white noise is independent of bandwidth until the critical band is exceeded, whereupon the loudness begins to increase. Their procedure was straightforward. They presented a band of filtered white noise and a comparison tone alternately through a single earphone. The subject adjusted the intensity of the tone until the tone and the noise sounded equally loud. The overall SPL of the noise was held constant; only the bandwidth was varied from judgment to judgment. (Zwicker and Feldtkeller did not report the number of subjects or the amount of variability; probably only a few, well-trained subjects were used and the variability was small.) Figure 2 shows what happens to the loudness of a band of noise when its width is increased. These curves are for bands centered at 1000 cps, which was the geometric mean of the two half-power points. At all the SPLs tested, from 30 to 80 db., the loudness of the noise remains constant and the curve is flat up to a bandwidth of about 160 cps, whereupon the loudness begins to increase. Within the critical band, the noises are as loud as a tone of equal intensity, having the same frequency as the center of the band. Functions similar in shape to those in Figure 2 were generated for bands centered at

500, 2000, and 4000 cps. The bandwidth at which loudness begins to increase defines the critical band for loudness, which was found to have approximately the same values as had been measured for threshold, two-tone masking, and phase sensitivity (see Figure 1).

Zwicker and Feldtkeller studied continuous spectra, i.e., noises that have energy at every frequency between the cutoff points. Bauch (1956) studied line spectra, i.e., sounds that have energy at two or more separate frequencies. He measured the loudness of three-tone complexes, produced by amplitude modulation, as a function of the difference, ΔF, in cps between the lowest and highest components of the complex. Bauch obtained the same results with three-tone complexes centered at various frequencies as Zwicker and Feldtkeller had obtained with bands of noise. For values of ΔF less than a critical band, loudness is constant except when ΔF is so small that beats are heard. The loudness begins to increase as a function of ΔF only when ΔF exceeds the critical band.

At the time that the critical band was being mapped out in Germany at the Technischen Hochschule Stuttgart (Bauch, 1956; Gässler, 1954; Zwicker, 1952, 1954; Zwicker & Feldtkeller, 1955) some of us at the Psycho-Acoustic Laboratory at Harvard were puzzled by our failure to find an increase in the loudness of a four-tone complex as a function of ΔF. We had assumed that loudness summation begins as soon as ΔF is increased. We were, however, studying four-tone complexes whose ΔFs were smaller than a critical band. When reports of the critical band came from Germany, our results began to make sense and, indeed, agreed

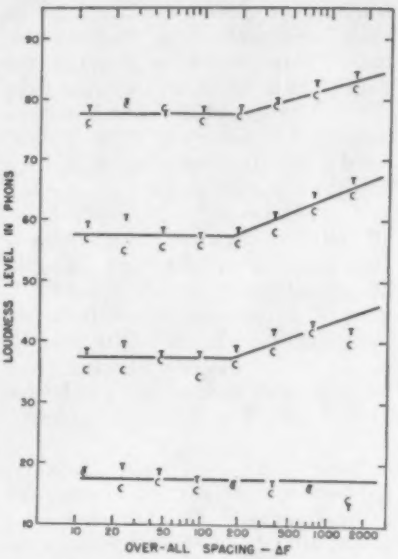

FIG. 3. The dependence of the loudness of a four-tone complex, centered at 1000 cps, on spacing and level. (Each point represents the median of two judgments by each of 10 listeners. The symbol T means the comparison tone was adjusted; C means the complex was adjusted. This figure is from Zwicker, Flottorp and Stevens—1957, p. 550.) (Reproduced with permission of the *Journal of the Acoustical Society of America*)

well with those being obtained across the sea. The experiments were continued at Harvard by S. S. Stevens with G. Flottorp from Norway and E. Zwicker from Germany (Zwicker et al., 1957). Four-tone complexes and bands of white noise, at various center frequencies and various SPLs, were studied. In these experiments, 16 to 22 untrained subjects sometimes adjusted the complex sound and sometimes adjusted the comparison until the two were equally loud. Figure 3 shows a typical set of results, those for four-tone complexes centered at 1000 cps. Each point is the median of 20 loudness matches. Although the

subjects were somewhat variable in their judgments, the medians are orderly and the lines through the data show a break at approximately the same value of ΔF that had been measured in Germany. The critical band made the transatlantic journey safely and invariantly.

Another investigation carried out at Harvard (Scharf, 1959a) showed that at low levels, between 5 and 35 db. above threshold, where the loudness of a complex sound increases more slowly with bandwidth than at higher levels, the critical band must be exceeded before loudness begins to change as a function of bandwidth.

Niese (1960), in Dresden, has also studied loudness summation and the critical band. He presented the sound stimuli not only through earphones (as in all the previous experiments) but also through a loudspeaker in a free field, i.e., in an anechoic room where sounds are almost completely absorbed by specially constructed walls. The results for free-field listening are similar to those for earphone listening; the loudness of a band of white noise begins to increase with bandwidth when the critical band is exceeded. Niese found, however, that the loudness did not continue to increase indefinitely with bandwidth, but increased about 8 db. and then remained constant for bandwidths greater than 1000 to 5000 cps depending upon the center frequency. It may be that the loudness did not increase further because the available energy was spread to very low and very high frequencies which contributed little to the total loudness.

In other experiments, Niese (1960) tested the assumption that loudness summation is a peripheral process occurring independently in each ear. In one procedure, a band of white noise was divided in half at its center frequency; the upper half was presented through an earphone to one ear and the lower half to the other ear. The loudness of the noise in both ears did not begin to increase with bandwidth until the *overall* width exceeded a value approximately *twice* the critical band, i.e., until the noise in each ear was wider than a single critical band. In a second procedure, two narrow bands, each 100 cycles wide, were first presented together to one ear and later separately to each ear. When presented together to a single ear, the loudness of the two bands increased with the frequency separation between them. When, on the other hand, one band was presented to each ear, the loudness did not increase with the frequency separation, no matter how great it was. The loudness did not increase because the band of noise presented to each ear was never wider than a critical band; it was always 100 cycles wide. Loudness summation thus seems to depend only upon the distribution of energy in one ear, suggesting that summation takes place not at some higher level in the auditory system where nerve impulses from the two ears join, but at the periphery, probably in the inner ear.

Still another aspect of loudness summation has been recently investigated (Scharf, 1959b). The results indicate that the loudness of a complex sound remains essentially unchanged when only the number of components in the complex is varied. The loudness of the complex increases with ΔF when ΔF is greater than a critical band, but at any given value of ΔF the loudness is approximately invariant with the number of com-

ponents, provided the overall sound pressure remains invariant.

The several experiments in loudness summation, along with those on threshold, two-tone masking, and phase sensitivity provide a firm body of evidence for the critical band. There remains, however, the question of the role of the critical band in the masking of pure tones by white noise.

Masking Bands

Although the empirical measures of the critical band are quite recent, the concept of a critical band was expounded some 20 years ago by Fletcher (1940) when he hypothesized that: (*a*) a pure tone that is masked by a white noise is in effect masked only by a narrow band of frequencies surrounding the tone, and (*b*) the intensity of the part of the band that does the masking is equal to the intensity of the tone.

Fletcher (1940) presented some preliminary experimental results to support his thesis, but the projected full-scale experiment has apparently not been reported. Nonetheless the concept of a critical band has become important in theories about masking. Moreover, the acceptance of Fletcher's hypotheses permits the calculation of values for the masking band from the measurement of the masking of pure tones by white noise (Hawkins & Stevens, 1950). The calculated values for the masking band turn out to be about two-and-one-half times smaller than the empirical values for the critical band, as measured in experiments on loudness, two-tone masking, etc. This discrepancy, however, may be resolved either by a modification of Fletcher's second hypothesis, or, better, by direct measurements of the masking band. Let us turn first to the indirect measurements of the masking band and the assumptions underlying them.

Indirect Measures of the Masking Band

If both Fletcher's hypotheses about the existence of a masking band and about the equality of the intensities of the tone and noise are accepted, it is possible to calculate the size of the masking band from the masked thresholds for pure tones in white noise. Only one empirical operation is necessary. The threshold for a tone is measured in the presence of a white noise. From the intensity of the just-masked tone and the intensity of the masking noise, it is fairly simple to calculate how large a band within the noise contains the same energy as the tone. The width of this band is, *by definition*, the masking band. Its width is calculated by taking the ratio of the intensity of the tone to the intensity per cycle of the noise. (Since a white noise contains all audible frequencies at equal intensity, the intensity per cycle is uniform throughout.) For example, Hawkins and Stevens (1950) found that the ratio between the intensity of a 1000-cycle tone (at its masked threshold) and the intensity per cycle of the masking noise is 63:1 or 18 db. Since the intensity in each one-cycle band of noise is 1/63 the intensity of the masked tone, a band of frequencies 63 cps wide will have an *overall* intensity equal to that of the tone. Therefore, according to the second hypothesis, the masking band is taken to be 63 cps wide for a tone of 1000 cps. Values for the masking band that are calculated in the foregoing manner will be called "critical ratios," as suggested by S. S. Stevens (see Zwicker et al., 1957).

Hawkins and Stevens measured the masked thresholds at many frequencies from 100 to 9000 cps in the presence of white noise at levels from 20 to 90 db. They found that the ratio of the intensity of a just-masked tone to the intensity per cycle of the masking noise remains constant at all noise levels except the very lowest. In other words, the critical ratio does not change as a function of the level of the masking noise. The critical ratio is, however, different at different center frequencies, as shown in Figure 1. The results of these experiments agree with similar measurements that Fletcher and Munson (1937) had made of the critical ratio for tones masked by a uniform masking noise.

Bilger and Hirsh (1956) also calculated critical ratios from masking data obtained with bands of white noise 250 mels wide. (The mel is a unit of pitch.) The substitution of a 250-mel band, which is about five times as wide as the critical ratios measured by Hawkins and Stevens, is consistent with the assumption that the energy outside the masking band contributes nothing to the masking effect. If this, Fletcher's fundamental assumption, is true the critical ratio should be the same in both experiments. The results of the two independent experiments were, in fact, in close agreement.

In all these experiments the calculated value of the critical ratio depends upon the measured value of the masked threshold which may not be very reliable. Blackwell (1953) has shown, for example, that the value obtained for a threshold depends upon the psychophysical method employed in its measurement. The congruence of the results of the several experiments tends, however, to negate this criticism. Using the reported threshold measurements, we can modify Fletcher's second asumption so that the masking band has the same values as the critical band.

Instead of assuming, quite arbitrarily, that the intensities of the masked tone and of the masking band are equal, we can just as well assume that the intensity of the masking band is two-and-one-half times as great as that of the masked tone. Over most of the frequency range, this simple modification of Fletcher's second hypothesis yields values for the masking band that are equal to the measured values of the critical band. A simple modification succeeds because, as Figure 1 shows, except for very low frequencies, the critical band and the critical ratio are the same functions of center frequency. Since this new assumption is ad hoc and arbitrary, it will probably have little appeal. What we need is a more direct and straightforward type of evidence of the existence of the masking band.

Direct Measures of the Masking Band

The direct measurement of the masking band requires the sampling of the masked threshold for tones in the presence of bands of noise of different widths. If a masking band exists, the tone should become more difficult to detect as the bandwidth of the noise is increased up to the value of the masking band. Increasing the bandwith beyond the masking band should not raise the threshold for the tone any further. (In such experiments, energy is added to the noise as the bandwidth is increased, unlike experiments on loudness summation where a constant amount of noise energy is spread over a wider frequency range in order to increase the bandwidth.) Direct measure-

ments of this type have been reported by Fletcher (1940), Hamilton (1957), and Schafer, Gales, Shewmaker, and Thompson (1950). Some of the recent experiments suggest that the masking band is larger than the critical ratio and may approximate the critical band as measured for other auditory phenomena.

In the first and most famous of these experiments, Fletcher (1940) measured the threshold for tones of seven different frequencies ranging from 125 to 8000 cps in the presence of bands of noise of various widths. No information about subjects, apparatus, or procedure was given. The results of this admittedly preliminary experiment provided some evidence for the masking-band hypothesis; the masked threshold tended first to increase and then to remain constant as the bandwidth of the masking noise was increased. The results seemed also to justify the assumption that, within the masking band, the intensity of the noise and the just-masked tone are equal: a band of noise, 30 cps wide, just masked a tone lying at its center frequency and having the same intensity. Precise determinations of the width of the masking band were not possible, however, because the data were highly variable and only a few bandwidths had been sampled. Of bandwidths having values in the vicinity of those for the masking band, only one, 200 cps wide, was adequately sampled. Nevertheless, relying heavily upon the assumption that the masking band and the just-masked tone are equally intense and upon the threshold measurements made in the presence of wide-band noise, Fletcher suggested values for the width of the masking band. These values, which Fletcher cautioned might be wrong by a factor of two, turned out to be approximately the same as the critical ratios calculated in 1950 by Hawkins and Stevens (see Figure 1). This similarity is not surprising, for the values recommended by Fletcher were, in effect, critical ratios. While suggestive, Fletcher's results provided neither conclusive support for his hypotheses nor a solid basis for the direct measurement of the width of the masking band.

Hamilton's (1957) more recent work provides a direct and precise measure of the masking band. Measuring the masked threshold for an 800-cycle tone in the presence of bands of noise that were centered at 800 cps and that varied in width from 19 to 1100 cps, he found that up to a bandwidth of 145 cps the masked threshold increased as the width of the masking noise increased. Beyond 145 cps the threshold remained constant, indicating that the masking band at 800 cps is 145 cps wide. The critical band measured in four other types of experiment is also about 145 cycles wide at 800 cps (see Figure 1). This coincidence of values is remarkable in view of the variability inherent in these experiments and Hamilton's apparent unfamiliarity with the other measures of the critical band.

A second important result in Hamilton's experiment shows that the difference (the signal/noise ratio) between the intensity of the 800-cycle tone at its masked threshold and the overall intensity of the masking noise is not constant, even when the width of the masking noise is less than a critical band. The signal/noise ratio decreases from about 0 db. for a band 30 cps wide to almost −4 db. for the critical width of 145 cps. (Hamilton reports similar results by Bauman, Dieter, Lieberman, and Finney, 1953.) Fletcher had also found that

a band 30 cps wide just masks a tone at its center when the signal/noise ratio is 0 db., i.e., when the intensities of the tone and the noise are equal. This equality at a width of 30 cps suggested that at the critical bandwidth also, the tone and noise have the same intensity. Hamilton showed, however, that at the critical bandwidth the signal/noise ratio is not the same as at 30 cps. Accordingly, Fletcher's threshold measurements for a tone in a 30-cps-wide band of noise probably lend no support to the critical-ratio hypothesis; they are, however, consistent with critical-band values for the masking band.

Although Hamilton studied only one frequency, his results provide valuable information because they are orderly and self-consistent. Probably the use of a forced-choice procedure with well-trained subjects contributed to the preciseness of the results. In contrast, Schafer et al. (1950) report a more extensive experiment whose results are difficult to interpret. They measured the masked threshold for tones in three frequency regions as a function of the bandwidth of the surrounding noise. Instead of the usual white noise, they used bands of synthetic noise composed of tones one cycle apart. Preliminary experiments indicated no important difference between these bands of synthetic noise and bands of white noise. Twenty-five subjects served in the main experiments in which a random method of limits was used to measure the masked threshold for a tone that had been matched in pitch to the masking noise. The results suggest the presence of a masking band, but since no sharp change in the masked threshold was observed as the bandwidth was increased, the width of the masking band can be estimated only approximately. In the three frequency regions that were tested, the results suggest a masking band that is larger than that given by the critical ratio, and one that could well be as large as a critical band.

Schafer et al. (1950) interpreted their results to indicate no change in the signal/noise ratio within the masking band. Hamilton (1957), on the other hand, did find a small but consistent change in the signal/noise ratio within the masking band. Since, however, Schafer's observers were too variable to permit a precise measurement of changes in the signal/noise ratio, the small difference between the results of the two experiments is probably not significant. There is also some question about what Schafer et al. measured. Their use of a tone "matched in pitch to the masking noise" may account for some of the disparity between their results and Hamilton's.

These two experiments, by Hamilton and by Schafer, seem to be the only direct tests of the masking-band hypothesis since Fletcher's original attempt. One related experiment (Webster, Miller, Thompson, & Davenport, 1952) deserves mention. A white noise with octave gaps was used to mask tones at frequencies corresponding to those in and near the gaps. The measurements of the masked thresholds seem to suggest that Fletcher's values for the masking bands are too small.

The lack of extensive tests of the masking-band hypothesis prevents a definitive statement about the validity of the hypothesis, and even less may be said about the size of the bands. Nevertheless the net impression one obtains from the literature is that a masking band does exist and

that it may well be the same width as the critical band.[5]

Other Correlates of the Critical Band

We have seen that the function relating the critical band to the frequency at the center of the band is derived from four types of experiment and that the width of the masking band may be the same as that of the critical band. Of interest, also, is the resemblance that the critical-band function bears to several other functions of frequency: the place of maximal displacement on the basilar membrane, the difference limen for frequency, and the mel scale of subjective pitch. These similarities have been noted elsewhere with respect to the critical band (Zwicker et al., 1957) and also with respect to the critical ratio (Fletcher, 1940, 1953; von Békésy & Rosenblith, 1951).

Perhaps the most interesting fact about the critical band is that it seems to correspond to a constant distance of about 1.3 millimeters along the basilar membrane. The first line in Figure 4 is a slightly idealized schematization of the frequency representation on the basilar membrane. The second line shows that 24 or 25 critical bands may be represented by equal-sized segments

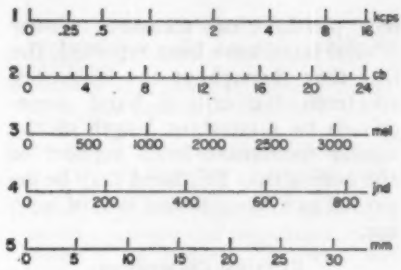

FIG. 4. Representation on the basilar membrane of (1) frequency in kilocycles, (2) critical bands, (3) pitch (Stevens & Volkmann, 1940), (4) just noticeable differences for frequency, the fifth line marks off distance in millimeters on the basilar membrane. (This figure is adapted from the book, *Das Ohr als Nachrichtenempfänger*, by Feldtkeller and Zwicker—1956, p. 60.) (Adapted with permission of S. Hirzel Verlag)

of the membrane. The boundaries of the critical bands are not fixed, of course, since a critical band may take shape around any frequency.

The mel and the jnd for frequency also correspond to constant distances on the basilar membrane (see the third and fourth lines in Figure 4). It is, therefore, not surprising that the critical-band function looks very much like the functions for the mel scale and the jnd scale. Measured in mels, the size of the critical band varies little, from 100 mels at low center frequencies to 180 mels at high frequencies. The mel scale is not accurate enough, however, to distinguish 100 from 180 mels at opposite ends of the scale, so that the pitch range of the critical band may, in fact, be fairly constant, perhaps approximating 150 mels.

The width of the critical band on the basilar membrane is determined from the map relating the frequency of pure tones to the position of maximal stimulation on the membrane (von Békésy, 1949). Although no di-

[5] Since the preparation of this article, Greenwood (1960) has reported an extensive study that confirms the suggestion that there is a masking band and that it is the same size as the critical band. Greenwood measured the threshold for pure tones presented in bands of white noise. He varied not only the width of the bands of noise around a given center frequency, but also the sensation level of the noise and the frequency of the masked tone. Investigating bands of noise in five regions of the spectrum, he found consistent evidence for the existence of a fairly sharp masking band approximately the same size as the critical band.

rect physiological measures of the critical band have been reported, the fact that throughout the frequency spectrum the critical band corresponds to a constant length of the basilar membrane lends support to the notion that this band may be regarded as a fundamental unit of hearing.

Future Prospects

With the experimental basis for the critical band reasonably well established, investigators are beginning to consider the relevance of the critical band to the loudness of pure tones, to temporal integration, to deafness, to speech perception, and to other auditory processes.

Zwicker (1956, 1958), for example, has argued that the loudness of an intense pure tone is a composite loudness because the displacement of the basilar membrane is spread over many critical bands. Zwicker assumes that the "loudnesses" corresponding to these critical bands summate to give the total loudness of the tone. Similar assumptions underlie Zwicker's (1958) system for the objective calculation of the loudness of a complex noise. The loudness of a noise is assumed to equal the sum of the individual loudnesses of the component critical bands after allowance for mutual masking effects among the bands.

Other investigators are studying temporal integration for short tone pulses (cf. Plomp & Bouman, 1959). Since short tone pulses are in effect multicomponent complexes whose bandwidth varies with time, the integration of energy at threshold would be expected to occur within the critical band.

Clinical use of the critical band has been attempted by deBoer (1960) in the diagnosis of hearing loss. His results suggest that the critical-band mechanism may be disturbed in certain kinds of deafness. The related problem of individual differences for the critical band has remained essentially uninvestigated except for some observations by Niese (1960) and indications from earlier data (e.g., Gässler, 1954) that the size of the critical band may vary from person to person, just as thresholds do.

Although no answers have yet come forth, phoneticists are beginning to ask about the role of the critical band in the perception of speech. Musicians may soon add their problems. The quest has begun in earnest. Now that a fundamental unit of hearing has been identified, it remains to discover its role in all the many processes called hearing.

REFERENCES

Bauch, H. Die Bedeutung der Frequenzgruppe für die Lautheit von Klängen. *Acustica*, 1956, **6**, 40–45.

Bauman, R. C., Dieter, C. L., Lieberman, W. J., & Finney, W. J. The effects of very narrow band filtering on the aural recognition of pulsed signals in noise backgrounds. *J. Acoust. Soc. Amer.*, 1953, **25**, 190. (Abstract)

Bilger, R. C., & Hirsh, I. J. Masking of tones by bands of noise. *J. Acoust. Soc. Amer.*, 1956, **28**, 623–630.

Blackwell, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bull. Engrg. Res. Inst., U. Mich.*, 1953, No. 36.

deBoer, E. Measurement of critical bandwidth in cases of perception deafness. *Proc. 3rd Int. Congr. Acoustics.* Amsterdam: Elsevier, 1960.

Feldtkeller, R. Ueber die Zerlegung des Schallspektrum in Frequenzgruppen durch das Gehör. *Elektrophys. Rdsch.*, 1955, **9**, 387.

Feldtkeller, R., & Zwicker, E. *Das Ohr als Nachrichtenempfänger.* Stuttgart: Hirzel, 1956.

FLETCHER, H. Auditory patterns. *Rev. mod. Phys.*, 1940, **12**, 47–65.

FLETCHER, H. *Speech and hearing in communication.* New York: Van Nostrand, 1953.

FLETCHER, H., & MUNSON, W. A. Relation between loudness and masking. *J. Acoust. Soc. Amer.*, 1937, **9**, 1–10.

GARNER, W. R. The effect of frequency spectrum on temporal integration of energy in the ear. *J. Acoust. Soc. Amer.*, 1947, **19**, 808–815.

GÄSSLER, G. Ueber die Hörschwelle für Schallereignisse mit verschieden breitem Frequenzspektrum. *Acustica*, 1954, **4**, 408–414.

GREENWOOD, D. D. Auditory masking and the critical band. Unpublished doctoral dissertation, Harvard University, 1960.

HAMILTON, P. M. Noise masked thresholds as a function of tonal duration and masking noise band width. *J. Acoust. Soc. Amer.*, 1957, **29**, 506–511.

HAWKINS, J. E., & STEVENS, S. S. The masking of pure tones and of speech by white noise. *J. Acoust. Soc. Amer.*, 1950, **22**, 6–13.

NIESE, H. Subjektive Messung der Lautstärke von Bandpassräuschen. *Hochfrequenztech. Elektroakust.*, 1960, **69**(1), 17.

PLOMP, R., & BOUMAN, M. A. Relation between hearing threshold and duration for tone pulses. *J. Acoust. Soc. Amer.*, 1959, **31**, 749–758.

SCHAFER, T. H., GALES, R. S., SHEWMAKER, C. A., & THOMPSON, P. O. Frequency selectivity of the ear as determined by masking experiments. *J. Acoust. Soc. Amer.*, 1950, **22**, 490–497.

SCHARF, B. Critical bands and the loudness of complex sounds near threshold. *J. Acoust. Soc. Amer.*, 1959, **31**, 365–370. (a)

SCHARF, B. Loudness of complex sounds as a function of the number of components. *J. Acoust. Soc. Amer.*, 1959, **31**, 783–785. (b)

STEVENS, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

STEVENS, S. S., & DAVIS, H. *Hearing: Its psychology and physiology.* New York: Wiley, 1938.

STEVENS, S. S., & VOLKMANN, J. The relation of pitch to frequency: A revised scale. *Amer. J. Psychol.*, 1940, **53**, 329–353.

VON BÉKÉSY, G. The vibration of the cochlear partition in anatomical preparations and in models of the inner ear. *J. Acoust. Soc. Amer.*, 1949, **21**, 233–245.

VON BÉKÉSY, G., & ROSENBLITH, W. A. The mechanical properties of the ear. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1075–1115.

WEBSTER, J. C., MILLER, P. H., THOMPSON, P. O., & DAVENPORT, E. W. The masking and pitch shifts of pure tones near abrupt changes in a thermal noise spectrum. *J. Acoust. Soc. Amer.*, 1952, **24**, 147–152.

ZWICKER, E. Die Grenzen der Hörbarkeit der Amplitudenmodulation and der Frequenzmodulation eines Tones. *Acustica*, 1952, **2**, 125–133.

ZWICKER, E. Die Verdeckung von Schmalbandgeräuschen durch Sinustöne. *Acustica*, 1954, **4**, 415–420.

ZWICKER, E. Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. *Acustica*, 1956, **6**, 365–381.

ZWICKER, E. Ueber psychologische und methodische Grundlagen der Lautheit. *Akust. Beih.*, 1958, **1**, 237–258.

ZWICKER, E., & FELDTKELLER, R. Ueber die Lautstärke von gleichförmigen Geräuschen. *Acustica*, 1955, **5**, 303–316.

ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. Critical band width in loudness summation. *J. Acoust. Soc. Amer.*, 1957, **29**, 548–557.

(Received January 7, 1960)

Psychological Bulletin
1961, Vol. 58, No. 3, 218–233

PERSEVERATIVE NEURAL PROCESSES AND CONSOLIDATION OF THE MEMORY TRACE[1]

STEPHEN E. GLICKMAN
Northwestern University

For a short period between the turn of the century and the first world war, theories of perseveration figured prominently in attempts to understand many of the newly discovered phenomena of learning and forgetting. Although the exact lines of speculation varied from one writer to the next, in general, a neural fixation process was assumed to continue after the organism was no longer confronted with the stimuli to be learned. This fixation process was deemed crucial to efficient retention and interference with perseveration was presumed to have an adverse effect on an organism's ability to remember stimuli to which it had been exposed.

The first clear statement of such a consolidation theory is generally attributed to Müller and Pilzecker (1900). In order to account for the existence of retroactive inhibition, Müller and Pilzecker postulated the existence of a neural perseverative process, subject to external interference and requisite to the consolidation of the memory trace for recently acquired material. Although knowledge of the physiology of brain function was still quite limited, Müller and Pilzecker nevertheless attempted to be as precise as possible regarding the neural locus of perseveration. They rejected the notion that perseveration was in any way analogous to sense organ processes such as those believed to underlie the negative afterimage, on the grounds that these sensory processes were of too short duration. On the other hand, the perseveration which Müller and Pilzecker observed did appear to be similar to the repetitious or stereotyped behavior resulting from diseases of the subcortical motor centers. It was with these latter structures that Müller and Pilzecker associated perseverative activity.

Numerous other psychologists were concerned with perseveration theory during the early 1900s. Among these, DeCamp (1915) advanced what was probably the most detailed piece of pseudoneurological speculation:

> From the neurological standpoint, in the learning of a series of syllables, we may assume that a certain group of synapses, nerve cells, nerve paths, centres, etc., are involved. Immediately after the learning process the after-discharge continues for a short time, tending to set associations between just learned syllables. Any mental activity engaged in during this after-discharge, involving or partially involving the same neurological group, tends, more or less, to block the after-discharge, and give rise to retroactive inhibition (p. 68).

Some years previous, Sherrington (1906) had described the phenomenon of afterdischarge in spinal reflexes and discussed the blockage of such discharges by subsequent stimuli. It is interesting to note that

[1] The preparation of this paper was made possible by a postdoctoral fellowhip held by the author at the Department of Physiology and Biophysics, University of Washington Medical School, during the summer of 1959, and supported by Grant 2-B5082 from the National Institute of Neurological Diseases and Blindness. The author is indebted to C. P. Duncan, S. M. Feldman, T. Kennedy, and D. Kimura for critically reading the manuscript.

this provided the theoretical model for DeCamp's view of perseverative processes in much the same manner as Sherringtonian physiology generally shaped the psychologists' conception of neural activity (see Hebb, 1951).

As a behavioral theory of retroactive inhibition, however, perseveration theory met with many difficulties, and was eventually replaced by the current concepts of associative interference (McGeoch & Irion, 1952; Osgood, 1953), although it continued to receive some limited support as a possible factor in forgetting (Woodworth, 1938). Ultimately, a perseveration theory, erected on the basis of inferences from behavior, was no longer viable once the behavioral observations were either shown to be false or explained more parsimoniously by other hypotheses. The rejuvenation of this theory awaited direct support from neurology.

Lashley (1918) once made the following comment on perseveration theory:

> If there is a gradual strengthening of associations during periods of nonpractice, there is implied a continuation of chemical changes within the nerve cells, initiated by the passage of a neural impulse through new channels and persisting for hours or even days without the influence of continued impulses. The experimental evidence upon which the belief in a gradual fixation of associations is based is far from convincing . . . it all can be explained equally well by other hypotheses and, in view of the extreme importance of the point for physiological explanation, we should be careful not to accept the assumption of a gradual setting of new functional connections until some real evidence is advanced to support it (pp. 363–364).

This healthy skepticism was certainly justified, although even at the time some physiological evidence was available to buttress perseveration theory.

Retrograde Amnesia

Shortly after the publication of Müller and Pilzecker's work, McDougall (1901) called attention to the applicability of their perseveration theory to the explanation of retrograde amnesia (RA) resulting from cerebral trauma. However, Burnham (1904) was apparently the first individual to extensively discuss the relationship between RA and perseverative "consolidation" amnesia. Burnham's paper involved an analysis of two cases of retrograde amnesia. Both of these subjects had sustained head injuries as the result of accidents and in both cases there was a loss of memory for events occurring during the period preceding the accident. As the result of his studies of these cases and of others cited by Ribot (1892), Burnham suggested that

> The fixing of an impression depends upon a physiological process. It takes time for an impression to become so fixed that it can be reproduced after a long interval; for it to become part of the permanent store of memory considerable time may be necessary. This we may suppose is not merely a process of making a permanent impression upon the nerve cells, but also a process of association, of organization of the new impressions with the old ones (p. 392).

He further speculated that: (*a*) the time required for this fixation process may vary with individuals and conditions; (*b*) shock produces its effects by arresting the fixation process in the nervous tissue; (*c*) such shock may be produced by great fatigue, excitement, unconsciousness, or narcotics; (*d*) RA is not all-or-none and the extent of the amnesia is relative to the amount of time elapsing before the fixation process is interrupted; and finally (*e*) that automatic activity is an important factor in fixing impressions although it may not

necessarily be directly observable in terms of movements.

These remarkable observations would appear to have been borne out by recent experiments in nearly every case and we can now advance these propositions with much more confidence.

During the first 4 decades of this century, the phenomenon of RA constituted the only direct physiological evidence for the existence of a neural fixation process. Early references to it are to be found in Ballard (1913), Pillsbury (1913), DeCamp (1915), and others. Although a complete review of this literature is beyond the scope of the present paper, it is perhaps worthwhile to examine the results of a comprehensive study by Russell and Nathan (1946). In a survey of 1,029 cases of head injury, only 133 were found to have experienced no RA whatsoever. Seven hundred and seven reported amnesia for events occurring from several seconds to 30 minutes preceding the injury, while 133 reported RA of more than 30-minutes duration. Records were unavailable with 56 patients in the sample. Russell and Nathan noted that the duration of RA is "in most cases a few moments only." Since the use of barbiturate hypnosis reduced the period of RA in only 6 of 40 cases, and produced no data suggestive of hysterical repressions, the authors conclude that loss of the material is due to a blocked perseveration process:

> It seems that the mere existence of the brain as a functioning organ must strengthen the roots of distant memories. The normal activity of the brain must steadily strengthen distant memories so that with the passage of time these become less vulnerable to the effects of head injury (p. 299).[2]

[2] Coons and Miller (1960) have recently called attention to the possibility of sampling artifacts confounding the consolidation interpretation of clinical observations of retrograde amnesia. Thus, they have pointed out that, if an injury produces a general decrement in memory, positive evidence for memory is more likely to be secured while examining the larger time samples involved in remote memories as compared to recent memories.

Experimentally induced RA has produced the best evidence for the existence of a consolidation process since the results would be predictable from perseveration theory, while the primary competing theory, the associative interference theory, has no explanation to offer. We will therefore turn now to a review of the various experimental procedures used to induce RA and the results obtained.

Electroconvulsive Shock

The introduction of electroshock therapy in 1937 provided both the impetus and the technical apparatus for the laboratory study of RA. Immediately after its introduction many practitioners observed that electroconvulsive shock (ECS) produced a temporary postshock amnesia which eventually shortened to a genuine RA for events immediately preceding the shock treatment. Zubin and Barrera (1941) were the first investigators to subject these observations to systematic study. They trained 10 patients in a series of paired associate lists to a criterion of two consecutive correct repetitions. Learning occurred either in the morning or evening, while the retention tests were given during the subsequent afternoon. The same subjects were used in control and experimental conditions, i.e., (*a*) with no shock intervening between learning and the retention test, and (*b*) with an ECS interpolated after the morning learning session. With no intervening shock there were significant savings between

learning and relearning, with an interpolated ECS there were no significant savings. A comparison between the effects of ECS on material learned the evening prior to shock with material learned the morning preceding shock indicated that recent material was more severely affected by ECS than remote material. The latter conclusion was based on rather small differences in savings scores and insufficient data are presented to permit adequate statistical evaluation. However, Flescher (1941), Williams (1950), and Cronholm and Molander (1958) have subsequently confirmed the substance of Zubin and Barrera's assertions. The various investigators using human subjects, although successfully employing ECS to interfere with memory, had not attempted to adequately define the time relations of such interference.

This critically important step was taken by Duncan (1949). Duncan's procedure involved training rats to avoid shock to the feet in a shuttlebox situation. A light, turned on 10 seconds prior to grid shock, served as the conditioned stimulus (CS). The animals received one trial per day for 18 days and records were kept of the number of successful avoidance responses. Nine groups of animals were used in the study. Rats in eight of these groups received an ECS after each day's trial, the trial-ECS interval ranging from 20 seconds to 14 hours. In the remaining group, the ear clips used for delivering the ECS were applied following each day's trial but no current was passed. The results clearly indicated a deleterious effect of ECS on performance, the magnitude of the effect decreasing as the trial-ECS interval increased to produce a negatively accelerated curve. This general finding has since been confirmed by Ransmeier (1953), Thompson and Dean (1955), and Leukel (1957). All of the findings are compatible with the view that a single ECS can produce deficits in retention if delivered within 15 to 60 minutes following a learning trial. Moreover, ECS induced immediately following the learning trial effectively obliterates nearly all retention of the "learned" response. The studies following Duncan's have employed different learning tasks. Leukel (1957) and Ransmeier (1953) used maze learning situations with the ECSs being delivered at varying posttrial intervals. Thompson and his collaborators have employed a visual discrimination learning task, with avoidance of grid shock as the motivating agent. In these latter studies (Thompson, 1957a; Thompson & Dean, 1955; Thompson & Pennington, 1957) a single ECS was administered at various intervals following a series of massed trials in the apparatus. As a result of these extensive experiments, it has been determined that ECS produces greater deficits in young than adult rats (Thompson, 1958a; Thompson, Haravey, Pennington, Smith, Gannon, & Stockwell, 1958). Further, rats suffering from anoxia induced brain damage (Pennington, 1958), show greater deficits resulting from a single ECS than intact control animals. Both the findings with respect to age and those relating to brain damage are compatible with Thompson and his co-workers' (1958) hypothesis that the extent of the deficit will be proportional to the number of cortical neurons available. Pennington (1958) has alternately suggested that the results obtained with brain damaged rats may be a function of a prolonged perseveration process in these animals.

Thompson and Pennington (1957)

have also found that the memory decrement produced by a single ECS was less after spaced trials than after massed trials. This result was expected from the point of view of a perseveration theory as a joint function of "firmer fixation of the memory trace owing to a longer duration of perseveration" and "the lessened intensity of perseveration at the end of training due to dissipation of perseverative activity."

Although the empirical result of interference with performance by postlearning ECS has not been questioned, the interpretation of the results is not quite as clear. The points to be discussed below actually raise questions of interpretation which apply not only to the ECS procedures but to other interpolated physiological procedures as well.

1. The most serious alternative to a consolidation interpretation of the ECS results has been offered by Miller and Coons (1955). These investigators trained rats to eat in a runway and then shocked them while eating there. Avoidance was measured by an increased latency of approach to the eating place. ECSs were delivered to the animals at varying intervals after shock to the mouth. Miller and Coons reasoned that any aversive qualities of the ECSs might be expected to produce increased avoidance. On the other hand, if the ECS really interrupted consolidation, the subjects would show the opposite behavior, namely, approaching the food without hesitation. In this experiment no evidence was found for an attenuation of the avoidance response by the ECS, leading the authors to argue that the retardation in learning observed by Duncan (1949) was simply a function of placing the rat in a conflict situation. In a more recent set of experiments, Coons and Miller (1960) have succeeded in opposing the conflict and consolidation interpretations in a double grid-box situation similar to that used by Duncan. Here again, their results indicate that ECS may not eliminate memory but merely induce anxiety or conflict which inhibits performance of the response in question. They further buttress their contentions regarding the fear inducing qualities of ECS with observations on increased defecation, urination, and weight loss in those animals for whom the performance of an otherwise rewarded response is followed by an ECS. In both of their studies, the ECS apparently summated with the grid shock to produce a result which significantly favored a conflict as opposed to a perseveration interpretation. Observations of Gallinek (1956) suggest that analogous anxiety builds up in human beings during the course of electroshock therapy. Such an interpretation is logically possible for both the avoidance situations used by Duncan and by Thompson and Dean, and the maze learning situations used by Leukel and by Ransmeier and Gerard. The standard control for this has been to employ groups receiving painful but nonconvulsive shocks. In these cases (Duncan, 1949; Leukel, 1957; Ransmeier & Gerard, 1954) it has been found that (*a*) the decrements produced by the painful but nonconvulsive shocks are not nearly as severe as those produced by ECS at comparable intervals, and (*b*) the posttrial interval during which painful shocks produced their effect was always much shorter than that during which significant decrements could be produced by ECS. These latter control results would seem to

indicate that the ECS results are due to more than just conflict. It might be argued, however, that the ECSs are sufficiently more painful or unpleasant than the leg or tail shocks to account for the greater deficits produced by the former. Reference to human subjects would suggest that this is not the case. Patients do not necessarily report pain as an accompaniment of a properly delivered ECS (Stainbrook, 1948). In view of this, and considering that there is no consistent experimental evidence for punishment obliterating verbal material (Rapaport, 1942), it seems unlikely that the deficits observed in humans following ECS (or any cerebral trauma) can be explained purely in conflict terms. Finally, in regard to the animal literature, it seems reasonable to point out that Miller and Coons delivered a series of ECSs on successive days, whereas Thompson and his collaborators eliminated a persistently rewarded response with a single ECS.

In order to explain Thompson's results in conflict terms, one would have to assume a delay of reinforcement gradient lasting at least 60 minutes, and the build up of a significant amount of fear following a single ECS. Such assumptions, although possible, would not be easy to support at the present time. Clearly, however, other workers should carry out experiments utilizing designs similar to those employed by Miller and Coons, i.e., opposing the consolidation and conflict interpretations. The writer has used such a procedure in an experiment involving direct stimulation of the brain (Glickman, 1958) and this could easily be adapted for ECS. Moreover, the one-trial learning situation employed in this latter experiment would permit the use of a single ECS and enable an exceedingly accurate estimate of the trial-ECS interval.[3]

2. An alternate interpretation of the ECS results is also possible in those studies employing food reward. As Kohn (1951) and Berkun, Kessen, and Miller (1952) have shown, the rewarding properties of food are derived in part from stimulation of receptors within the mouth, and in part from actions within the stomach. ECS delivered shortly after a learning trial might act to prevent the perception of the feedback from the stomach and thereby cut down on the reinforcing properties of the food. In view of the relatively minor contribution of these stomach receptors, particularly in the early stages of learning, such effects are probably insignificant in the studies of Ransmeier (1953) and Leukel (1957).

3. A question has arisen about distinguishing between ECS effects on a time-limited consolidation process and the more generalized memory deficits which have been observed to follow a series of ECSs (see Stainbrook, 1946, for review). In particular, Worchel and Gentry (1950) have suggested that Duncan's (1949) finding of a limited period following learning when an ECS will be effective is a result of his failure to used massed ECSs. On the basis of some T maze data of their own, Worchel and Gentry argue that Duncan might

[3] Since this article went to press, there have been two reports of experiments in which the conflict and consolidation interpretations of "forgetting" have been opposed in one-trial learning situations. In both of these cases, in which the introduction of various chemical agents served as the interpolated procedure, the results favored a consolidation interpretation of the effects (Essman & Jarvik, 1960; Pearlman, Sharpless, & Jarvik, 1961).

have considerably extended the duration of time during which ECS would produce a deficit if if he had given a series of ECSs. Worchel and Gentry's results do not contradict the general finding that it is easier to disrupt learning in the period immediately following exposure to the learning situation. However, at the present time, the ECS data are compatible with the notion that the strengthening of memory traces is a continuous one throughout the life of the organism. For example, Brady (1951, 1952) has found evidence of spontaneous growth in the strength of a conditioned emotional response during a period of 90 days. On the basis of current evidence, one might expect that the interval following a learning trial, during which time interference with retention can be produced, is a direct function of the degree of physiological severity of the interpolated procedure.

Ultimately, there is probably some practical limit on the time interval between learning and ECS during which selective effects on retention can be produced. In addition, the effects of a series of ECSs delivered many hours or days after learning are often apparently temporary (Stainbrook, 1946). Brady (1951) reported that a series of ECSs suppressed a conditioned emotional response (CER) for a period of a month, although the habit reappeared spontaneously at the end of that time. It has also been found that the effects of a series of ECSs may be selective for emotional responses (Geller, Sidman, & Brady, 1955). On the basis of the accumulated data, it seems reasonable to suggest that ECS may affect performance in a number of ways including: (*a*) a temporary suppressor action involving those cerebral structures mediating pain or anxiety responses (such a mode of action would explain the proactive effects noted by Poschel, 1957, and Carson, 1957, on avoidance conditioning) and (*b*) a direct action on the neural circuits involved in memory which, if the learning-ECS interval is brief enough and the treatment sufficiently severe, may permanently erase the effects of such learning.

Anoxia

Hayes (1953) demonstrated equivalent retroactive effects of anoxia and ECS on maze learning in rats. He used a distributed practice procedure and administered the experimental treatment one hour after each trial. The experimental rats showed similar retardation in learning when their acquisition curves were compared with normal control animals. Hayes reports that histological examination of the brains produced no clear evidence of brain damage for any of the animals. Ransmeier and Gerard (1954) have also reported disturbances in maze learning resulting from anoxia, the magnitude of the disturbance decreasing "along characteristic curves with increasing intervals between training and experimental procedures."

Using a discrimination learning procedure, Thompson and Pryer (1956) showed that anoxia, produced by placing rats in a decompression chamber during the postlearning period, could lead to decrements in retention analogous to those produced by ECS. In a later study, Thompson (1957a) found that a 10-minute exposure to a simulated 30,000-foot altitude produced deficits equivalent to those resulting from ECS, although exposure to a 20,000-foot altitude did not produce such severe effects. Finally, Thompson

(1957a) has also reported that when an ECS was given 30 seconds post-training, a subsequent 10-minute exposure to a simulated 30,000-foot altitude did not produce an additional deficit.

Temperature

A number of investigators have studied the effects of postlearning temperature on retention. In most of the earlier work (French, 1942; Hunter, 1932; Jones, 1943) the aim was to reduce the activity of the experimental group and thereby reduce retroactive inhibition. Considered in the light of the ECS literature, these studies are not immediately relevant to the present review because of the prolonged interval between the learning trials and the achievement of the desired temperature change.

In the most recent studies of Cerf and Otis (1957) and Ransmeier and Gerard (1954) it appears that temperature may have some effect on processes related to consolidation. The former investigators gave goldfish 10 massed trials in an avoidance situation using a shifting light as the CS. At varying intervals following the trials, (0 minute, 15 minutes, 60 minutes, or 4 hours) the body temperatures of different groups of 15 to 19 subjects were raised briefly to a point sufficient to induce heat narcosis (36.5°–37.0° C). In retention tests carried out the next day, the criterion of five consecutive correct responses in 10 trials was met by only 10.5% of the group narcotized immediately after learning, while 56.2% of the subjects paralyzed 4 hours following learning met the same criterion. The remaining two groups occupied intermediate positions. Fifty percent of a group of untreated control subjects also met the above criterion. Thus, the temperature induced narcosis produced much the same effect in the goldfish that ECS and anoxia have been found to produce in rodents. Ransmeier and Gerard (1954) did not find any evidence of retroactive effects of lowered body temperatures on retention of a maze habit in the hamster. Gerard (1955) has reported, however, that lowering the body temperature will apparently prolong the period during which an ECS may produce severe deficits. Thus, "hamsters kept cool between learning and electroshock show as great a disruption of learning at an interval of one hour as warm ones do at an interval of fifteen minutes." Evidently, temperatures sufficient to impair spontaneous activity in the brain as indicated by the EEG will not act directly to block consolidation, although they may slow down the chemical processes involved in the fixation of the trace.

Fay (1940) has reported RA in human subjects for events occurring while the patients were refrigerated, i.e., when the body temperature fell below 33.3° C. Under these circumstances, the subjects could respond to questions and carry on a conversation, although interrogation after the refrigeration procedure showed a loss of memory for the entire interchange. Such deficits could be explained in terms of an impairment of activity in those structures responsible for the consolidation process. However, alternative explanations are also possible.

Anesthesia

Leukel (1957) has reported that sodium pentothal injected intraperitoneally (IP) after each learning trial impaired acquisition in a maze in experimental rats when their time or error scores were compared with any of three control groups. Subjects

in the three control groups received either: an IP injection of water following each trial, an IP injection of pentothal 30 minutes following each trial, or no injection. The scores did not differ among these latter groups. Leukel interpreted his results in terms of interruption of consolidation of the memory trace in those subjects receiving pentothal one minute after each trial.

On the other hand, Russell and Hunter (1937) and Ransmeier and Gerard (1954) have not found deficits in retention to result from postlearning barbiturate anesthesia. There are numerous differences in procedure, however, which might account for this discrepancy. For example, Russell and Hunter (1937) administered sodium amytal subcutaneously after giving their experimental subjects five massed trials in a maze. They observed no effects of the injection on subsequent retention of the maze. However, the subcutaneous route of injection undoubtedly prolonged the time before the drug took effect (in comparison with the IP route used by Leukel). In addition, the massed trials procedure used by Russell and Hunter resulted in a longer interval between learning and anesthesia than the Leukel procedure of injecting one minute following each trial.

Ransmeier and Gerard (1954) and D. Kimura and S.E. Glickman (unpublished) failed to find retention deficits as the result of anesthetizing hamsters or rats with ether following maze learning trials, or electric shock in an avoidance learning situation, respectively. These results suggest that the apparent effectiveness of barbiturates, as opposed to ether, in blocking consolidation may be due to secondary effects of the former on blood chemistry or blood pressure rather than direct synaptic interference. Barbiturate anesthetics produce many more severe blood changes than ether including reductions in blood pressure and blood sugar level (Kohn, 1950).

If anesthetics can be shown to exert reliable retroactive effects on learning, they may eventually prove useful in the localization of the neural structures crucial to consolidation. Techniques have recently been developed which permit the delivery of small quantities of various drugs to restricted sites within the brain of a "behaving" animal (Fisher, 1956; Olds & Olds, 1958). Utilizing such techniques, it should be possible to selectively and temporarily block activity in various cerebral structures during the period immediately following exposure to the learning situation and thereby determine which structures, if any, are crucial to the consolidation process.

Brain Stimulation

Mahut (1958), Glickman (1958), and Thompson (1958b) have reported retroactive effects of brain stimulation on learning. The stimulation was accomplished with chronically implanted electrodes which permit the animal freedom of movement in the learning situation, but enable the experimenter to deliver a small electric current to particular sites within the CNS at any chosen time. This technique enables much more specific delimitation of the structures involved in the presumed fixation process than, for example, ECS or anoxia. However, in the studies carried out thus far, there are numerous factors which serve to complicate comparisons among the studies, as well as to rule out any simple "consolidation" interpretation of the results.

Mahut (1958) tested the effects of stimulation of the nonspecific thalamic nuclei on the performance of rats in a Hebb-Williams maze. Brief bursts of 60-cycle, sine wave, 0.25-volt stimulation were delivered through implanted electrodes while the rat was eating in the goal box. Such stimulation produced poorer performance in the maze, when the error scores of these "thalamic" rats are compared with those of rats receiving either no stimulation or similar stimulation of the midbrain tegmentum. The possibility exists in this study that the effects of stimulation were not retroactive but contemporary, i.e., interfered with the animals' registration of the food reward. This might be clarified by a parametric investigation of the time interval between learning trial and stimulation, following the design of the ECS studies (Duncan, 1949; Thompson & Dean, 1955).

Glickman (1958) examined the effects of stimulation of the midbrain portion of the arousal system on the acquisition of an avoidance habit in the rat. Three 20-second bursts of stimulation, at considerably higher voltages than those used by Mahut (1958), were delivered immediately following shock to the mouth while the subjects were eating at a distinctive metal food spout. In retention tests carried out the following day, the animals who had received reticular stimulation after mouth-shock showed less avoidance of the spout (more eating behavior) than control animals not receiving brain stimulation. The interpretation of this study is also complicated due to the particular characteristics of the Hudson (1950) one-trial learning apparatus which evidently lead to a portion of the avoidance response being learned in the postshock period. Hudson has reported that the visual scanning which the animal engages in during the postshock period will reinforce the avoidance response. Thus, it is conceivable that the reticular stimulation could have simply interfered with an ongoing visual process rather than retroactively interfering with previous learning.

Thompson (1958b), in an ingeniously designed study which permits him to use each animal in a variety of experimental conditions, has reported interference with the performance of cats in an alternation task as the result of intracranial stimulation. This effect was achieved with bilateral stimulation of the caudate nucleus following each trial in a modified Wisconsin General Test Apparatus. Similar stimulation of the midbrain tegmentum did not produce the retroactive effect, although it did interfere with performance when the stimulation was delivered either before or during a given trial. In this case, the interpretation of the retroactive disruptive effects of caudate stimulation is complicated by the possible reinforcing properties of this stimulation. Brady, Boren, Conrad, and Sidman (1957) have reported positively reinforcing consequences of caudate stimulation in the cat. It seems plausible that stimulation in this region, following a particular response, would favor repetition of that response and might act in opposition to any alternation habit. Such an explanation might be an alternative to postulating interference with a perseveratory process. Since it is possible to check on the rewarding properties of electrical stimulation, using a self-stimulation situation such as that used by Olds and Milner (1954), this factor could be easily controlled in future studies. In regard to the lack of effect of teg-

mental stimulation, this may be explicable in terms of the extensive functional localization of reinforcement pathways which appears to exist in that region (Glickman, 1960; Olds & Peretz, 1959). Olds[4] has suggested that the interference produced by intracranial stimulation in learning situations may be directly related to the reinforcing qualities of the stimulation.

There are numerous studies demonstrating interference with learning as a result of intracranial stimulation (see Zeigler, 1957, for review). However, most of these are not directly interpretable in terms of retroactive interference because the stimulation is delivered during the actual performance of the task. Nevertheless, as Thompson (1958b) suggests, interference with consolidation may be at least a partial explanation of the deficits observed by Rosvold and Delgado (1956) coincident with caudate stimulation. Similarly, Burns and Mogenson (1958) and Burns and Stackhouse (1959) have reported deficits in the acquisition of a bar pressing habit in the Skinner Box resulting from a cortical stimulation. As Burns and Stackhouse note, these results are compatible with a perseveration hypothesis.

Physiological Substrate of Consolidation

Stellar (1957) has pointed out that physiological data have recently accumulated which tend to support the existence of a system within the brain responsible for the permanent fixation of memory traces. Milner and Penfield (1955) and Scoville and Milner (1957) have reported cases of temporal lobe ablation in man which produced severe impairment of the ability to acquire new material postoperatively, although preoperatively acquired material was retained. Although the crucial structures have not yet been definitely localized, the hippocampus and amygdala appear to be directly involved. Along similar lines Brady, Schreiner, Geller, and Kling (1954) found interfering effects of amygdalectomy on the acquisition of an avoidance response in cats, although the same lesions produced in cats which had already acquired the habit led to no disturbance in performance. The anatomical and physiological data suggest numerous pathways through which these relatively primitive temporal lobe structures could exert widespread effects on the remainder of the brain (Adey, Merrillees, & Sunderland, 1956; Green & Adey, 1956). For example, the continued action of these temporal lobe regions may be necessary to the proper regulation of firing in the nonspecific arousal system, which in turn apparently exerts considerable influence on cortical activity (Magoun, 1958).

The existence of structures within the brain which are crucial to the fixation of memory traces is not restricted to the vertebrate orders. Boycott and Young (1950) have identified a cerebral structure (the vertical lobe) requisite for fixation of visual memory in the octopus, and apparently homologous in function to the temporal lobe structures found in the higher vertebrates. Thus, removal of the vertical lobe drastically impairs the ability of the animal either to acquire a new visual discrimination habit (motivated by a combination of food and electric shock), or to retain such a habit for any length of time following training. The nervous system of the octopus differs widely from the vertebrate

[4] J. Olds, personal communication, 1959.

nervous system. However, the appearance of a specialized fixation mechanism in both invertebrates and vertebrates suggests that there is some evolutionary utility in a dual process underlying memory function.

At a more molecular level, the most widespread hypothesis concerning the substrate of consolidation predicates its dependence on reverberatory circuits. This idea has its origins in the anatomical demonstrations of Lorente de No (1938) and has been subscribed to in varying forms by Hebb (1949), Young (1953), and Gerard (1955). The basic supposition is that reverberatory activity maintains the memory until the permanent changes underlying fixation of the trace have been completed. This dual process hypothesis of memory fixation has the advantage of explaining why interference with neural activity immediately after "learning" blocks retention while similar procedures instituted at a later time do not. One group of studies which may be directly relevant to the reverberatory circuit hypothesis of consolidation has been carried out by B. D. Burns and his co-workers (Burns, 1954, 1958). Burns has developed a technique which allows the isolation of small areas of cortex from the remainder of the brain, while leaving the blood supply to the area relatively unaffected. He has extensively studied the electrical activity of these isolated slabs in response to direct electrical stimulation. Interestingly enough, he has found: that a single train of pulses can initiate bursts of activity in one of these preparations lasting for 30 minutes or more; that such bursts of activity can be blocked by a subsequently applied electrical stimulus; that such activity becomes easier to evoke with repeated applications of the stimulus; and that the burst activity is apparently due, in part to reverberatory activity among groups of neurons, and in part to differential rates of depolarization within various segments of individual neurons. These first three observations certainly coincide with what one would expect if such a process underlay consolidation. However, it is necessary to be cautious in generalizing from the type of activity observed in these special preparations to that occurring in the intact brain. Burns (1958) himself has rejected these preparations as a general model for memory on the grounds that such circuits would be too susceptible to external interference. However, this is one aspect of the data which makes Burns' findings so attractive as a model of the first phase of a dual process theory of memory, susceptibility of learned material to interference providing the main behavioral evidence for the existence of a consolidation process.

Finally, moving to a still more molecular analysis of the problem, it is reasonable to inquire about the specific changes which might be produced by some sort of perseverative process. Nearly all investigators have at this level proposed some sort of growth process or chemical change at the synapse. In this respect, our ideas have changed little from those of 1929 when Lashley wrote:

> We have today an almost universal acceptance of the theory that learning consists of modification of the resistance of specific synapses within definite conduction units of the nervous system.

After expressing numerous reservations about the adequacy of this assumption, Lashley concluded by noting that:

> The synapse is, physiologically, a convention to describe the polarity of conduction in the

nervous system of higher animals, together with some similarities of function in the central nervous system and neuromuscular junction. That these functions are due to the action of the intercellular membranes has not been directly demonstrated (p. 127).

Here again, recent neurophysiological progress tempers Lashley's skepticism. The synapse is no longer a "convention" but a point-at-able structure which can be photographed and studied with the electron microscope (Palay, 1956). Further, as Lloyd (1949) and Eccles (1953) have shown, the rapid firing of impulses across synaptic junctions can result in increased excitability of these synapses for periods lasting from minutes to hours. There is general agreement that this increased excitability results from the firing of presynaptic fibers, although it is not yet clear whether this is in turn due to an actual change in the dimensions of the synaptic knobs as suggested by Eccles (1953, 1957) or if an alternate explanation, e.g., Lloyd (1949), may suffice. Eccles (1953) has proposed this phenomenon of posttetanic potentiation as a general model for conditioning and memory. Such a proposal meets with many difficulties (Malmo, 1954). However, there is no question that a person ascribing learning to changes in synaptic excitability could do so with more confidence today than was possible 30 years ago.

Conclusions

In the opinion of the writer, the over-all weight of evidence certainly favors the existence of some mechanism of consolidation (in spite of the fact that alternative explanations are possible for many of the experiments which supposedly support the existence of such a process). Furthermore, the application of available physiological procedures appears to offer a promising approach to defining the structures involved in the fixation of memory traces. The most severe problems presented thus far have occurred as the result of confounds in the behavioral test situations employed, rather than through some defect in the modes of physiological interference. These problems are not insoluble, however, and an attempt was made to indicate this in the text of the paper.

As a final point, the material reviewed suggests the possibility that pseudoneurological speculation, resulting from strictly behavioral observation, can result in productive physiological research—when the speculation is shrewdly conceived. Moreover, the physiologist would appear to have already begun to repay this debt by suggesting purely behavioral studies or new interpretations of behavioral data. The studies demonstrating interfering effects of visual stimulation interpolated immediately after visual discrimination learning (Thompson, 1957b; Thompson & Bryant, 1955) are examples of such physiologically influenced "behavioral" investigations. Along similar lines, Walker's (1958) reinterpretation of reaction decrement, spontaneous alternation data, in terms of mechanisms serving to protect consolidation, appears to be equally sensitive to current physiological research.

REFERENCES

Adey, W. R., Merrillees, N. C. R., & Sunderland, S. The entorhinal area: Behavioural, evoked potential, and histological studies of its interrelationships with brain-stem regions. *Brain*, 1956, 79, 414–439.

BALLARD, P. B. Oblivescence and reminiscence. *Brit. J. Psychol., Monogr. Suppl.*, 1913, **1**, No. 2.

BERKUN, M. M., KESSEN, MARION L., & MILLER, N. E. Hunger-reducing effects of food by stomach fistula versus food by mouth measured by a consummatory response. *J. comp. physiol. Psychol.*, 1952, **45**, 550–554.

BOYCOTT, B. B., & YOUNG, J. Z. The comparative study of learning. *Symp. Soc. Exp. Biol.*, 1950, **4**, 432.

BRADY, J. V. The effect of electroconvulsive shock on a conditioned emotional response: The permanence of the effect. *J. comp. physiol. Psychol.*, 1951, **44**, 507–511.

BRADY, J. V. The effect of electroconvulsive shock on a conditioned emotional response: The significance of the interval between the emotional conditioning and the electroconvulsive shock. *J. comp. physiol. Psychol.*, 1952, **45**, 9–13.

BRADY, J. V., BOREN, J. J., CONRAD, D., & SIDMAN, M. The effect of food and water deprivation upon intracranial self-stimulation. *J. comp. physiol. Psychol.*, 1957, **50**, 134–137.

BRADY, J. V., SCHREINER, L., GELLER, I., & KLING, A. Subcortical mechanisms in emotional behavior: The effect of rhinencephalic injury upon the acquisition and retention of a conditioned avoidance response in cats. *J. comp. physiol. Psychol.*, 1954, **47**, 179–186.

BURNHAM, W. H. Retroactive amnesia: Illustrative cases and a tentative explanation. *Amer. J. Psychol.*, 1903, **14**, 382–396.

BURNS, B. D. The production of afterbursts in isolated unanesthetized cerebral cortex. *J. Physiol.*, 1954, **125**, 427–446.

BURNS, B. D. *The mammalian cerebral cortex.* London: Arnold, 1958.

BURNS, N. M., & MOGENSON, G. Effects of cortical stimulation on habit acquisition. *Canad. J. Psychol.*, 1958, **12**, 77–82.

BURNS, N. M., & STACKHOUSE, S. P. Effects of cortical stimulation on habit acquisition: Confirmatory data. *Amer. Psychologist*, 1959, **14**, 427. (Abstract)

CARSON, R. C. The effect of electroconvulsive shock on a learned avoidance response. *J. comp. physiol. Psychol.*, 1957, **50**, 125–129.

CERF, J. A., & OTIS, L. S. Heat narcosis and its effect on retention of a learned behavior in the goldfish. *Federation Proc.*, 1957, **16**, 20–21. (Abstract)

COONS, E. E., & MILLER, N. E. Conflict versus consolidation of memory traces to explain "retrograde amnesia" produced by ECS. *J. comp. physiol. Psychol.*, 1960, **53**, 524–531.

CRONHOLM, B., & MOLANDER, L. Influence of an interpolated ECS on retention of memory material. *U. Stockholm Psychol. Lab. Rep.*, 1958, No. 61.

DECAMP, J. E. A study of retroactive inhibition. *Psychol. Monogr.*, 1915, **19**(4, Whole No. 84).

DUNCAN, C. P. The retroactive effect of electroshock on learning. *J. comp. physiol. Psychol.*, 1949, **42**, 32–44.

ECCLES, J. C. *The neurophysiological basis of mind.* Oxford: Clarendon, 1953.

ECCLES, J. C. *The physiology of nerve cells.* Baltimore: Johns Hopkins, 1957.

ESSMAN, W. B., & JARVIK, M. E. The retrograde effect of ether anesthesia on a conditioned avoidance response in mice. *Amer. Psychologist*, 1960, **15**, 498. (Abstract)

FAY, T. Observations on prolonged human refrigeration. *NY State J. Med.*, 1940, **40**, 1351–1354.

FISHER, A. E. Maternal and sexual behavior induced by intracranial stimulation. *Science*, 1956, **124**, 228–229.

FLESCHER, D. L'amnesia retrograda dopo l'ettroshock: Contributo allo studio della patogenesi delle amnesia in genere. *Schweiz. Arch. Neurol. Psychiat.*, 1941, **48**, 1–28.

FRENCH, J. W. The effect of temperature on the retention of a maze habit in fish. *J. exp. Psychol.*, 1942, **31**, 79–87.

GALLINEK, A. Fear and anxiety in the course of electroshock therapy. *Amer. J. Psychiat.*, 1956, **113**, 428–434.

GELLER, I., SIDMAN, M., & BRADY, J. V. The effect of electroconvulsive shock on a conditioned emotional response: A control for acquisition recency. *J. comp. physiol. Psychol.*, 1955, **48**, 130–131.

GERARD, R. W. Biological root of psychiatry. *Science*, 1955, **122**, 225–230.

GLICKMAN, S. E. Deficits in avoidance learning produced by stimulation of the ascending reticular formation. *Canad. J. Psychol.*, 1958, **12**, 97–102.

GLICKMAN, S. E. Reinforcing properties of arousal. *J. comp. physiol. Psychol.*, 1960, **53**, 68–71.

GREEN, J. D., & ADEY, W. R. Electrophysiological studies of hippocampal connections and excitability. *EEG clin. Neurophysiol.*, 1956, **8**, 245–262.

HAYES, K. J. Anoxic and convulsive amnesia in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 216–217.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. The role of neurological ideas in psychology. *J. Pers.*, 1951, **20**, 39–55.

HUDSON, B. B. One-trial learning in the domestic rat. *Genet. psychol. Monogr.*, 1950, **41**, 99–145.

HUNTER, W. S. The effect of inactivity produced by cold upon learning and retention in the cockroach, *Blatella Germanica*. *J. genet. Psychol.*, 1932, **41**, 253–266.

JONES, M. R. The effect of hypothermia on retention. *J. comp. physiol. Psychol.*, 1943, **35**, 311–316.

KOHN, H. I. Changes in blood of the rat during ether and barbiturate anaesthesia. *Amer. J. Physiol.*, 1950, **160**, 277–284.

KOHN, M. Satiation of hunger from food injected directly into the stomach versus food injected by mouth. *J. comp. physiol. Psychol.*, 1951, **44**, 412–422.

LASHLEY, K. S. A simple maze: With data on the relation of the distribution of practice to the rate of learning. *Psychobiology*, 1918, **1**, 353–367.

LASHLEY, K. S. *Brain mechanisms and intelligence*. Chicago: Univer. Chicago Press, 1929.

LEUKEL, F. P. The effect of ECS and pentothal anaesthesia on maze learning and retention. *J. comp. physiol. Psychol.*, 1957, **50**, 300–306.

LLOYD, D. P. C. Post-tetanic saturation of response in monosynaptic reflex pathways of the spinal cord. *J. gen. Physiol.*, 1949, **33**, 147–170.

LORENTE DE NO, R. Analysis of the activity of the chains of internuncial neurons. *J. Neurophysiol.*, 1938, **1**, 207–244.

MCDOUGALL, W. Experimentelle Beiträge zur Lehre von Gedächtniss: Von G. E. Müller und A. Pilzecker. *Mind*, 1901, **10**, 388–394.

MCGEOCH, J. A. & IRION, A. L. *The psychology of human learning*. (2nd ed.) New York: Longmans, Green, 1952.

MAGOUN, H. W. *The waking brain*. Springfield, Ill.: Charles C Thomas, 1958.

MAHUT, HELEN. Discussion. In H. H. Jasper, L. D. Proctor, R. S. Knighton, W. C. Noshay, & R. T. Costello (Eds.), *Reticular formation of the brain*. Boston: Little, Brown, 1958.

MALMO, R. B. Eccles' neurophysiological model of the conditioned reflex. *Canad. J. Psychol.*, 1954, **8**, 125–129.

MILLER, N. E., & COONS, E. E. Conflict versus consolidation of memory to explain "retrograde amnesia" produced by ECS. *Amer. Psychologist*, 1955, **10**, 394–395. (Abstract)

MILNER, BRENDA, & PENFIELD, W. The effect of hippocampal lesions on recent memory. *Trans. Amer. Neurol. Ass.*, 1955, 42–48.

MÜLLER, G. E., & PILZECKER, A. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Z. Psychol.*, 1900, Suppl. No. 1.

OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. comp. physiol. Psychol.*, 1954, **47**, 419–427.

OLDS, J., & OLDS, M. E. Positive reinforcement produced by stimulating hypothalamus with iproniazid and other compounds. *Science*, 1958, **127**, 1175–1176.

OLDS, J., & PERETZ, B. Relation of arousal and motivational effects in reticular activating system. *Federation Proc.*, 1959, **18**, 116. (Abstract)

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

PALAY, S. L. Synapses in the central nervous system. *J. biophysic. biochem. Cytol., Suppl.*, 1956, **2**, 193–202.

PEARLMAN, C. A., JR., SHARPLESS, S. K., & JARVIK, M. E. Retrograde amnesia produced by anesthetic and convulsant agents. *J. comp. physiol. Psychol.*, 1961, **54**, 109–112.

PENNINGTON, D. F. The effects of ECS on retention of a discrimination habit in rats subjected to anoxia. *J. comp. physiol. Psychol.*, 1958, **51**, 687–689.

PILLSBURY, W. B. *The essentials of psychology*. New York: Macmillan, 1913.

POSCHEL, P. B. H. Proactive and retroactive effects of electroconvulsive shock on approach-avoidance conflict. *J. comp. physiol. Psychol.*, 1957, **50**, 392–396.

RANSMEIER, R. E. The effects of convulsion, hypoxia, hypothermia and anesthesia on retention in the master. Unpublished doctoral dissertation, University of Chicago, 1953.

RANSMEIER, R. E., & GERARD, R. W. Effects of temperature, convulsion and metabolic factor on rodent memory and EEG. *Amer. J. Physiol.*, 1954, **179**, 663–664. (Abstract)

RAPAPORT, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.

RIBOT, T. Memory. In D. H. Tuke (Ed.), *A dictionary of psychological medicine*. Philadelphia: Blakiston, 1892.

ROSVOLD, H. E., & DELGADO, J. M. R. The effect on delayed-alternation test performance of stimulating or destroying electrically structures within the frontal lobes of the monkey's brain. *J. comp. physiol. Psychol.*, 1956, **49**, 365–372.

RUSSELL, R. W., & HUNTER, W. S. The effects

of inactivity produced by sodium amytal on the retention of the maze habit in the albino rat. *J. exp. Psychol.*, 1937, **20**, 426–436.

RUSSELL, W. R., & NATHAN, P. W. Traumatic amnesia. *Brain*, 1946, **69**, 280–300.

SCOVILLE, W. B., & MILNER, BRENDA. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiat.*, 1957, **20**, 11–21.

SHERRINGTON, C. S. *The integrative action of the nervous system.* New York: Scribner, 1906.

STAINBROOK, E. Shock therapy: Psychologic theory and research. *Psychol. Bull.*, 1946, **43**, 21–60.

STAINBROOK, E. The effects of electrically induced convulsions on animal behavior. *J. Pers.*, 1948, **17**, 2–8.

STELLAR, E. Physiological psychology. *Annu. Rev. Psychol.*, 1957, **8**, 415–436.

THOMPSON, R. The comparative effects of ECS and anoxia on memory. *J. comp. physiol. Psychol.*, 1957, **50**, 397–400. (a)

THOMPSON, R. Retroactive effect of interpolated visual stimulation. *Psychol. Rep.*, 1957, **3**, 183–188. (b)

THOMPSON, R. The effect of ECS on retention in young and adult rats. *J. comp. physiol. Psychol.*, 1958, **50**, 644–646. (a)

THOMPSON, R. The effect of intracranial stimulation on memory in cats. *J. comp. physiol. Psychol.*, 1958, **51**, 421–426. (b)

THOMPSON, R., & BRYANT, J. H. Memory as affected by activity of the relevant receptor. *Psychol. Rep.*, 1955, **1**, 393–400.

THOMPSON, R., & DEAN, W. A further study on the retroactive effect of ECS. *J. comp. physiol. Psychol.*, 1955, **48**, 488–491.

THOMPSON, R., HARAVEY, F., PENNINGTON, D. F., SMITH, J., JR., GANNON, D., & STOCKWELL, F. An analysis of the differential effects on ECS on memory in young and adult rats. *Canad. J. Psychol.*, 1958, **12**, 83–96.

THOMPSON, R., & PENNINGTON, D. F. Memory decrement produced by ECS as a function of the distribution of original learning. *J. comp. physiol. Psychol.*, 1957, **50**, 401–404.

THOMPSON, R., & PRYER, R. S. The effect of anoxia on the retention of a discrimination habit. *J. comp. physiol. Psychol.*, 1956, **49**, 297–300.

WALKER, E. L. Action decrement and its relation to learning. *Psychol. Rev.*, 1958, **65**, 129–142.

WILLIAMS, M. Memory studies in electric convulsion therapy. *J. Neurol. Neurosurg. Psychiat.*, 1950, **13**, 30–35.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

WORCHEL, P., & GENTRY, G. Electroconvulsive shock and memory: The effect of shocks administered in rapid succession. *Comp. psychol. Monogr.*, 1950, **20**, No. 104, 95–119.

YOUNG, J. Z. Discrimination and learning in octopus. In H. Von Foerster (Ed.), *Cybernetics: Transactions of the ninth conference.* New York: Josiah Macy, Jr. Foundation, 1953.

ZEIGLER, H. P. Electrical stimulation of the brain and the psychophysiology of learning and motivation. *Psychol. Bull.*, 1957, **54**, 363–382.

ZUBIN, J., & BARRERA, S. E. Effect of electric convulsive therapy on memory. *Proc. Soc. Exp. Biol. Med.*, 1941, **48**, 596–597. (Abstract)

(Received February 5, 1960)

Psychological Bulletin
1961, Vol. 58, No. 3, 234–255

THE ASSESSMENT OF ANXIETY BY PHYSIOLOGICAL BEHAVIORAL MEASURES[1]

BARCLAY MARTIN
University of Wisconsin

The term anxiety has enjoyed great popularity in the writings and researches of psychologists in the last decade, and procedures for measuring this hypothetical state have proliferated wildly. There is every indication that psychologists will continue to develop and employ measures of anxiety in many areas of research, especially in the rapidly expanding number of studies of psychotherapeutic process and change, in the already booming area of psychopharmacology, in studying the effects of anxiety on performance, and in attempts to assess such constructs as aggression anxiety or sex anxiety. It is the purpose of this paper to first impose some restrictions upon the definition of anxiety, and then to focus upon the problem of assessment by physiological-behavioral measures. No attempt will be made in this paper to review the research and evaluate the problems associated with assessing anxiety by self-report techniques.

One's theoretical approach to anxiety affects how one goes about measuring it; likewise results of attempts to assess anxiety should eventually modify and help refine the theoretical conception of anxiety. Thus the initial comments about the nature of anxiety should be considered as a rough formulation only, with both assessment procedures and theory modifying each other as investigation proceeds. It is recognized that this formulation, rough as it is, cannot include all that anxiety means to all people, and that accordingly to make this review manageable, it is necessary to delimit the concept.

[1] This paper was prepared in part while the author held a visiting appointment at the University of North Carolina. The author wishes to thank Earl Baughman and Leonard Berkowitz for their helpful suggestions.

A Conception of Anxiety

As a starting point it is proposed that the construct of anxiety be considered similar and perhaps identical to the reaction of fear, the neurophysiological bases for which are not completely known but would seem to especially involve the functions of the posterior hypothalamus and its effects upon the sympathetic nervous system, the adrenal medulla, and the pituitary-adrenocortical system. The brain stem reticular formation may also play a part in this reaction. It is recognized that this is undoubtedly an oversimplification of the complex and interacting neurophysiological mechanisms involved in fear. This reaction may be largely innate yet it is likely that as a result of learning or constitutional predisposition individuals tend to have variations in the manner in which the anxiety reaction is expressed.

It is further proposed that anxiety represents only one of many arousal states that can be differentiated from a more general state of activation as arousal becomes more intense. Thus the arousal that occurs when a person passes from a sleeping or very relaxed state to a waking, behaving state may be of a fairly generalized sort with no specialized affective or motivational reactions involved. However, as arousal becomes more intense, differentiation probably occurs and distinctive arousal states may emerge relating to such constructs as anxiety,

anger, hunger, sex, or other emotional or motivational states. Although it is possible that research will suggest the value of distinguishing anxiety from fear at the response level, or one kind of anxiety from another, it is perhaps best to demonstrate the utility of one construct of anxiety and its distinctiveness from other arousal states before adding unnecessarily to the number of theoretical constructs extant.

Anxiety also possesses the property of being highly learnable: that is, the hypothetical response becomes readily conditioned to stimuli that do not innately elicit the response. This characteristic renders difficult if not impossible any attempt to define anxiety on the basis of stimuli that elicit it, since the stimuli that elicit it will vary widely from person to person. An exception would be direct electrical stimulation of the brain (Miller, 1958), where the effective antecedent stimulus might be well defined.

As a consequence of the difficulty of approaching the construct of anxiety from the stimulus side in human subjects, the primary emphasis in this paper will be to review research relevant to the assessment of anxiety in terms of response patterns. The observable responses from which one might infer the strength of the anxiety reaction are of two basic types: physiological-behavioral responses and self-report responses. As previously mentioned, this paper will be primarily concerned with the first type of response.

In addition to the hypothetical anxiety state and its observable manifestations there are two other variables intimately related to anxiety which are kept conceptually distinct in the present view: namely, those stimuli (external or internal) which elicit the anxiety response, and those responses which have been learned because they reduce or avoid the anxiety response. From the point of view of measurement the stimuli that evoke anxiety become important only if one wants to know what situations or thoughts or feelings elicit anxiety. Thus the common distinction between anxiety and fear in terms of the latter being in response to a realistic danger and the former being a response to unrealistic or unknown threats is basically a stimulus defined difference and does not necessarily involve a difference in response.

There exists a possible source of confusion with respect to the responses that have been learned to reduce anxiety in that clinicians frequently infer anxiety on the basis of these "defenses" against anxiety as much as from direct expression of the anxiety itself. Again, from the point of view of theory as well as measurement it is preferable to keep these two variables distinct if possible. In fact, it would seem likely that when a person is making a successful "defensive" response, no anxiety is present. To the extent that this is so it would be misleading to infer the strength of the momentary anxiety level from the presence of learned anxiety reducing responses.

The Measurement of Anxiety

The foregoing theoretical analysis suggests that in spite of individual variations in response there might still be some pattern of physiological-behavioral responses associated with anxiety arousal that would be distinct from other patterns of response associated with other emotional or arousal states. Findings based primarily on physiological response patterns will be considered first followed by findings based primarily on behavioral response patterns. Two basic questions will be asked with respect to both the physiological and the behavioral evidence: (*a*) Does a dis-

tinctive pattern of responses emerge, tentatively identifiable as reflecting anxiety, that can be distinguished from other patterns associated with other arousal states, when the differing arousal states have been experimentally induced? (*b*) What is the nature of the intercorrelations among physiological or behavioral measures which have been obtained under the same experimental conditions, and is there any evidence of a distinguishable cluster of intercorrelated variables that might be tentatively identified as reflecting anxiety? The studies do not always lend themselves to a clear-cut analysis in these terms but these are the guiding questions being considered.

Physiological Measures: Experimental Comparisons

The studies of primary interest here are those in which an attempt was made to distinguish between two or more experimentally induced arousal states where one of these was considered to represent a fear or anxiety reaction. There are three studies that most closely follow this paradigm. Ax (1953) reports a study in which a variety of physiological measures were obtained from normals under conditions presented in counterbalanced order that were designed to elicit fear and anger, respectively. The fear condition was ingeniously contrived to make the subject think that the apparatus was faulty and that he was in real danger of receiving a severe, perhaps even fatal, electric shock. Anger was aroused by an obnoxious assistant who generally insulted and belittled the subject. Schachter (1957) repeated Ax' study using hypertensive, potential hypertensive, and normotensive subjects, and added a pain experience (cold pressor test) to the fear and anger situations. All subjects received the treatments in the same order: pain, fear, and anger. Lewinsohn (1956) obtained three physiological measures plus a measure of finger tremor on groups of normals, anxiety reaction patients, ulcer patients, and hypertensive patients subjected in counterbalanced order to the cold pressor test and a failure experience accompanied by criticism and electric shock. Another study that is highly relevant to the issue but which employed a somewhat different research strategy is that of Funkenstein, King, and Drolette (1957). After stressing their college student subjects they determined in a poststress interview whether a subject had tended to experience anger outwardly directed, anger inwardly directed, or anxiety. The scores obtained were limited to blood pressure and ballistocardiographic measures.

The results of these four studies are summarized in Table 1. Most scores in the Ax and Schachter studies represent difference scores between prestress resting level and the highest (or in some cases the lowest) level reached during stress. The scores in the Lewinsohn study represent differences in the mean during rest and the mean during stress, with the exception of the GSR score which represents the largest deflection during stress. All scores reported from the Funkenstein study are percentage changes from prestress levels.

In spite of some inconsistencies among the studies there does appear to be evidence for distinguishable response patterns that can be tentatively associated with the constructs of fear (anxiety) and anger. Diastolic blood pressure increased more for anger than fear in all three studies in which fear and anger states were thought to be aroused (significantly different from chance in two studies). Heart rate increased more in fear than anger in all three studies (significant in two). Maximum heart rate

TABLE 1

COMPARISON OF PHYSIOLOGICAL MEASURES ASSOCIATED WITH DIFFERENT EMOTIONAL AROUSAL STATES IN FOUR STUDIES

Measure	Ax		Schachter			Lewinsohn		Funkenstein	
	Fear	Anger	Fear	Anger	Pain	Fear	Pain	Fear	Anger-out
Systolic blood pressure	20.4	19.2	22.5	21.1	17.8			19.6%*	13.1%*
Diastolic blood pressure	14.5*	17.8*	13.7	14.5	11.8			9.7%	22.8%*
Heart rate (+)	30.3	25.8	18.7*	10.8*	0.3*	5.4*	0.9*	33.3%*	7.4%*
Heart rate (−)	4.0*	6.0*							
Cardiac output			6.7*	3.0*	−0.25*			61.9%*	−3.2%*
Peripheral resistance			−1.10*	0.04*	1.28*			−19.3%*	32.9%*
Hand temperature (−)	.045	.050	.036*	.030*	.024*				
Palmar conductance	14.8*	9.4*	−1.99**	−2.18**	−2.33**				
Largest deflection in stress, GSR						2.52	2.15		
No. GSRs	4.7*	11.6*							
Respiratory rate	6.0*	2.3*	2.8*	2.1*	0.7*				
Frontalis muscle tension	3.34*	4.35*	1.30	2.26	1.65				
No. muscle potential peaks	13.2*	10.5*							
Finger tremor						87	118		
Salivary output						−0.9	7.9		

* Schachter used the transformation, log $1/(R_1 - R_2)$, where R_1 = initial resistance and R_2 = lowest resistance during stress. The smallest negative number, −1.99, for fear accordingly refers to the largest decrease in resistance.
* Significant at the .05 level; for Schachter this is based on an overall analysis of variance for the three conditions.

decrease was significantly greater in anger than fear in the one study in which it was reported. Cardiac output increased significantly more in fear than anger in the two studies in which it was reported, and peripheral resistance decreased significantly more in fear than anger in both studies where it was reported. Palmar conductance increased significantly more in fear than anger in the two studies where it was reported. Number of discrete GSRs, however, was significantly higher in anger than fear in the one study where this was measured. Respiration rate increased significantly more in fear than anger in the two studies reporting this measure. Frontalis muscle tension increased more in fear than anger in the two studies measuring it (significant in one).

Another study that was not included in the tabular presentation provides additional support for the different heart rate responses associated with anxiety and anger. DiMascio, Boyd, and Greenblatt (1957) studied one psychotherapy patient over 11 interviews and found a correlation (rho) of .69 between average heart rate and amount of rated tension (anxiety?) in the interviews, and a correlation of −.37 between average heart rate and amount of rated antagonism in the interviews.

The two studies in Table 1 involving a painful experience, the cold pressor test, suggest that this arousal state may also be distinguishable from fear, although the differentiation of pain and anger is less clear. It is, of course, not possible to know from these results how specific these reactions might be to the cold pressor test as opposed to pain stimulation generally.

Funkenstein et al. (1957) propose a theory that may serve to provide some integration for these various findings. They suggest that the physiological reaction accompanying anger-out is a norepinephrine-like reaction and that accompanying anxiety is an epinephrine-like reaction. The physiological reactions accompanying injections of epinephrine and norepinephrine have been investigated by Goldenberg, Pines, Baldwin, Greene, and Roh (1948), Barcroft and Konzett (1949), DeLargy, Greenfield, McCorry, and Whelan (1950), Goldenberg (1951), Swan (1952), and Clemens (1957). In general it is found that epinephrine leads to increased palmar con-

ductance, systolic blood pressure, heart rate, cardiac output, forehead temperature, central nervous system stimulation, blood sugar level; and decreased diastolic blood pressure, peripheral resistance, hand temperature, and salivary output. Norepinephrine leads to increased systolic and diastolic blood pressure and peripheral resistance, no change or a slight decrease in heart rate and cardiac output, and only slight increases in central nervous system stimulation and blood sugar level.

It is generally thought that reactions associated with norepinephrine are more limited, possibly restricted to peripheral vasoconstriction resulting from secretion at the sympathetic nerve endings, than are the reactions to epinephrine. However, no studies were found in which the effects of injected norepinephrine upon a wide range of responses including palmar conductance, hand or finger temperature, respiration rate, salivary output, or muscle potentials were assessed. In terms of the measures that have been obtained under both kinds of hormonal injections (Barcroft & Konzett, 1949; DeLargy et al., 1950; Goldenberg et al., 1948), heart rate, diastolic blood pressure, cardiac output, and peripheral resistance appear to be the most discriminating. Neither cardiac output nor peripheral resistance is readily obtainable by direct measurement. Cardiac output is usually inferred from ballistocardiographic measures, and peripheral resistance is usually estimated by dividing mean arterial blood pressure by cardiac output.

Funkenstein et al. (1957) divided their subjects into subgroups on the basis of epinephrine-like, norepinephrine-like, and indeterminate reactions and found a highly significant relationship in the expected direction between these physiological reaction types and the tendency to respond by anger-out as opposed to anxiety. Schachter (1957) making use of a greater variety of physiological measures likewise computed an index of epinephrine- and norepinephrine-like reactions and found these indices to vary significantly as a function of the pain, anger, and fear conditions with pain showing the most norepinephrine-like reaction and fear the most epinephrine-like reaction with anger falling in between.

Although it would be premature to conceptualize the anxiety reaction as being entirely defined by the results of epinephrine secretion, the distinction between the epinephrine- and norepinephrine-like reactions may well be an important one for anxiety measurement. The secretion of epinephrine and norepinephrine from the adrenal medulla and the release of norepinephrine at the sympathetic nerve endings are all affected by sympathetic nervous system stimulation. The fact that these two hormones produce quite different reactions points up what has long been known: namely, that it is a great oversimplification to speak of sympathetic arousal as if it were a unitary function. Although the response pattern associated with experimentally induced anxiety conforms rather closely to the response pattern associated with epinephrine injection, the response pattern associated with anger is not as closely related to the responses produced by norepinephrine injection. Perhaps the distinction between anxiety and anger, at the humoral level, is one involving the relation of epinephrine to norepinephrine in which anxiety is associated with a purer epinephrine-like reaction and anger with a mixed pattern of epinephrine and norepinephrine responses.

There are other studies where one or two physiological measures have been obtained under conditions likely

to arouse anxiety. For example, Hickham, Cargill, and Golden (1948) found heart rate and cardiac output to increase substantially in medical students before what was considered to be an anxiety arousing situation, an oral examination, as compared to more relaxed conditions a month later. Likewise, Malmo, Boag, and Smith (1957) report increased heart rate in neurotic subjects after criticism as compared with decreased heart rate after praise. Although studies of this kind tend to be consistent with the previously described studies, they do not shed additional light on the question of whether some pattern of response related to anxiety can be differentiated from patterns of response associated with other kinds of arousal states.

Davis (1957), Davis and Buchwald (1957), and Davis, Buchwald, and Frankman (1955) also report evidence that different stimuli elicit distinctive autonomic response patterns. There is no reason to believe, however, that any of their stimuli, for example, pictures of nudes, landscapes, etc., were likely to evoke anxiety in many of their subjects. These studies do point to the possible subtleties in autonomic patterns associated with various kinds of stimulation or arousal states, and caution against any too ready acceptance of some particular pattern as being *the* anxiety or *the* anger pattern. All of the studies described thus far, though, are consistent with the possibility that some pattern of physiological measures may allow one to infer the magnitude of the hypothetical anxiety reaction differentially from other hypothetical states such as anger or pain.

Physiological Measures: Group Comparisons

There is a host of studies in which physiological measures are contrasted between normals and various clinical groups presumed to be in general more anxious than the normals. The studies that will be considered here are those involving patient groups in which the presence of manifest anxiety was reported to be a prominent part of the symptom picture; accordingly, much of the physiological research on such psychosomatic disorders as hypertension and peptic ulcer will not be summarized.

Sherman and Jost (1942) found 15 neurotic children to have *lower* resting level palmar conductance than 18 well adjusted children, but more resting level hand tremors, lower percentage of alpha rhythm in the EEG, and faster respiration rate than well adjusted children. No differences were found for heart rate or blood pressure. Although measures were taken in a series of seven conditions, the results described above appeared to represent differences in general level rather than different degrees of reaction to the various conditions. Jurko, Jost, and Hill (1952) obtained measures on 25 normals, 20 neurotics, and 10 schizophrenics (all adults) while administering the Rosenzweig P–F test, and found heart rate, respiration rate, and respiration variability higher in patient than normal groups before and during the test administration. A body movement score was highest for the schizophrenics and lowest for the normals. Palmar conductance was again found to be inconsistent with the general pattern, being highest for the normals and lowest for the schizophrenics before and during test administration. In neither of these two studies, however, was any attempt made to restrict the sample of neurotics to patients in which anxiety was the most prominent symptom.

GSR conditioning rate, on the other hand, has been found to be

faster in more anxious subjects (Bitterman & Holtzman, 1952; Schiff, Dougan, & Welch, 1949; Welch & Kubis, 1947).

White and Gildea (1937) found that patients in which anxiety was a prominent symptom showed greater heart rate increases to the cold pressor test than did normals. On the surface such a finding appears contradictory to the results of Schachter (1957) in which the physiological responses associated with the cold pressor test were clearly distinguishable from those associated with anxiety. White and Gildea, however, obtained measures during a rest period, during a brief anticipation period in which the experimenter moved the dish of ice water close to the subject, and during the immersion itself. For the normal group the average heart rates for these three periods were 75.7, 81.5, and 80.0, respectively; and for a group of anxiety neurotics 81.0, 90.0, and 95.5, respectively. Clearly, it was the anticipation of the experience that led to increased heart rate for the normals, not the pain experience itself. The anxious patients likewise showed their greatest increase during anticipation. These results suggest that anticipation of the cold pressor test is anxiety arousing, and might yield a different pattern of response, in normals at any rate, than the pain experience itself.

The above results of White and Gildea as well as the results of Schachter (1957) and Lewinsohn (1956) argue against the theoretical formulation of Mowrer (1939) that anxiety (fear) is the conditioned form of the pain reaction.

In the Lewinsohn (1956) study previously mentioned, resting level palmar conductance was highest for the anxiety reaction group and lowest for the ulcer group, with normals and hypertensives falling in between. Resting level salivary output was highest in the ulcer group, and lower and about the same for the other three groups. Somewhat surprisingly, resting level heart rate was lowest for the anxiety group. The change scores showed no particular tendency to be associated with the diagnostic groups. Wishner (1953) found resting level heart rate to be higher in 11 anxiety neurotics than in 10 normals and a tendency, not significant, for respiration rate to be faster in the neurotics. Funkenstein, Greenblatt, and Solomon (1951, 1952) conclude that patients with anxiety and depressive symptoms are manifesting a chronic epinephrine-like reaction, whereas patients with paranoid tendencies or who are otherwise directing their anger and blame upon the external world are manifesting a chronic norepinephrine-like reaction. Their conclusions are based primarily on the patients' reactions to the mecholyl test (Funkenstein, Greenblatt, & Solomon, 1950).

Malmo (1950, 1957) has summarized his research with respect to physiological measures found to discriminate between normals and patients with pathological degrees of anxiety. In his 1957 article he concludes that anxious patients show greater reactivity in many measures regardless of the kind of stress used. Thus, Malmo and Shagass (1949a) using a painful thermal stimulation of the forehead as their stress found anxiety neurotics and early schizophrenics to show more finger movements, greater neck muscle potentials, more head movements, more respiratory irregularities, and greater heart rate variability than normal controls. Percent change of the GSR showed no significant relationship. These results have been generally borne out in other studies using different stresses: Malmo, Shagass, and Davis (1951); Malmo, Shagass, Belanger, and Smith (1951). The results

of Malmo and Smith (1955) suggest frontalis muscle tension may be a more sensitive discriminator between normals and anxiety neurotics than forearm muscle tension.

Wenger (1948) using considerably larger *N*s than most investigators compared resting state physiological measures of 225 patients with the diagnosis of operational fatigue, 98 hospitalized psychoneurotics, and a normative group of 488 unselected preflight students in the Army Air Force. The 10 measures that significantly discriminated between the operational fatigue group and the normal group were salivary output, palmar conductance, systolic and diastolic blood pressures, sinus arrhythmia, heart period, sublingual temperature, finger temperature, respiration period, and tidal air mean. The operational fatigue group showed sympathetic dominance on all of the above measures except sublingual temperature. For 47 patients in the operational fatigue group Wenger obtained repeat measures on most variables at a later time when they were considered improved and ready to return to duty. Of the 20 variables tested only palmar conductance, heart period, and finger temperature showed significant changes, and these were all in the direction of lessened sympathetic arousal. The results with respect to the hospitalized psychoneurotics, although not yielding exact correspondence on specific measures, also showed a strong sympathetic dominance for this clinical group.

Gunderson (1953) obtained 12 resting state autonomic measures, selected on the basis of Wenger's previous work, on a sample of 110 early schizophrenics with an average length of hospitalization of about 2 years. Nine measures—salivary output, dermographic latency, dermographic persistence, systolic blood pressure, diastolic blood pressure, finger temperature, heart rate, respiration rate, and sublingual temperature—were significantly different from Wenger's normative group of aviation cadets, and with the exception of sublingual temperature all were in the direction of greater sympathetic arousal. Palmar conductance failed to discriminate and was, in fact, almost identical for the two groups. This schizophrenic sample also showed significantly greater sympathetic arousal in seven of these measures than Wenger's neurotic group. As Gunderson points out this indication of greater anxiety in the schizophrenic group may well not exist in more chronic patients. Gunderson also divided the schizophrenic subjects into those that had improved the most and least with shock therapy and found the most improved group to show less general sympathetic arousal as measured by Wenger's autonomic balance score, the conclusion being that improvement had been accompanied by a decreased arousal.

There are difficulties involved in comparing these studies in which anxiety is assumed to be present by virtue of a psychiatric diagnosis with those in which anxiety was produced experimentally. For example, if anger or annoyance does involve a distinctive arousal state and if such a state is present more often in some of these patient groups than in normals, a not unlikely assumption, then the pattern of mean scores may reflect a mixture of anxiety and anger as well as other arousal states. Nevertheless, many measures which belong to the epinephrine-like pattern of reaction are found to consistently discriminate, with an occasional exception, between anxious patients and normals. By and large it would appear that so-called resting state measures discriminate between the patients

and normals as well, and in some cases better, than do change scores associated with experimental stress. Some of the studies reporting change score results may be misleading since in most cases the patient groups start out with higher initial level scores. The high negative correlation between initial level and the magnitude of the change score that prevails for most autonomic measures might well obscure some real differences that would have emerged if this correlation had been partialed out by some procedure such as Lacey's (1956) autonomic lability score.

It is also possible that the particular pattern of autonomic responses associated with an immediate threat situation is different from the "steady state" pattern of more chronically elevated responses found in many psychiatric patients. It is interesting in this regard that Wenger (1957) in recent pattern analyses of the data in his various samples reports not only patterns of sympathetic and parasympathetic dominance but a pattern composed of a mixture of sympathetic and parasympathetic type of responses. This latter pattern, which Wenger calls the B pattern, consists of three sympathetic type tendencies, high heart rate, high systolic blood pressure, and low salivary output; and two characteristics of parasympathetic innervation or lack of sympathetic arousal, high finger temperature and low palmar conductance. The sympathetic pattern occurred more frequently in neurotic and schizophrenic samples than in the normal group, but not more frequently in the operational fatigue or a psychosomatic sample than in the normal group. The B pattern occurred more frequently in all of the four psychiatric groups than in the normal group. Perhaps this B pattern represents a more chronic result of psychological stress which could be distinguished from the anxiety state as presently conceived. Such an interpretation is consistent with the common clinical view that psychosomatic symptoms frequently serve an anxiety reducing function. It is also noteworthy that the findings of Sherman and Jost (1942) and Jurko, Jost, and Hill (1952) of low resting level palmar conductance in a pattern otherwise suggestive of sympathetic activation in neurotic patients is consistent with the existence of Wenger's B pattern.

To carry speculation a bit further in this area, it may be that there are systematic differences in response patterns as a function of the chronicity of the stress, as suggested by Selye (1950). Thus the pattern(s) of immediate change scores associated with discrete stimuli (electric shock or a threatening word) may be different from the pattern(s) of response associated with a stress of longer duration but still essentially temporary or situational (oral examination, the general situation in an electric shock experiment, or an appointment for a first psychotherapy hour), where the change scores would have to be based upon measures obtained at some more relaxed time. And both of the above kinds of patterns might differ from patterns of response resulting from stress continuing over months or years as would be the case with psychiatric patients. The distinctive characteristics of responses associated with the second as opposed to the first type of stress may result from humoral effects being added to the more direct and shorter latency effects of autonomic nervous system stimulation.

There have been several other approaches to the physiological assessment of anxiety employing measures less readily obtainable and also less amenable to continuous recording than most of the ones considered

above. Ulett, Gleser, Winokur, and Lawler (1953) and Shagass (1955b) report that the EEG of anxious patients can be more readily "driven" at higher frequencies than is the case for normals or less anxious patients. There was no tendency for the average undriven alpha frequency to be different for the groups. Shagass (1955a) further reports that changes in the driven EEG frequency correspond to changes in anxiety level for the same person measured at different times.

Sedation threshold is also reported by Shagass (1954) and Shagass and Naiman (1955) to be related to anxiety level in patients. Basowitz, Persky, Korchin, and Grinker (1955) find more hippuric acid in the urine of paratrooper trainees assessed to be anxious than those not anxious, and also more in anxiety neurotics than in normals.

Physiological Measures: Intercorrelations

On the basis of the research just summarized one might assume that many of the measures found to be related to experimentally induced or clinically assessed anxiety would show substantial intercorrelations. Research thus far gives little ground for optimism that these variables will correlate very highly, if at all. However, it should be pointed out that there are few researches that provide much direct evidence on the question: namely, correlations among changes in measures obtained under resting and a clearly fear or anxiety arousing situation. Ax (1953) intercorrelated the seven physiological change scores that significantly discriminated between the fear and anger conditions. The intercorrelations of these scores under the anger condition tended to be higher than for the fear condition. The correlations were for the most part insignificant for fear. Schachter (1957) did not report intercorrelations among his measures but did find significantly more variability among the measures under fear than anger. Lewinsohn (1956) likewise reported intercorrelations among his four variables for base level scores, for change scores to the cold pressor test, and for change scores to the failure-criticism condition. Only a few correlations were significant, probably not more than could have occurred by chance. Terry (1953) intercorrelated a number of physiological change scores associated with doing arithmetic problems under distracting noise conditions. The intercorrelations between different autonomic systems were very low and for the most part insignificant. Only measures of closely related functions, such as systolic and diastolic blood pressure, correlated to any degree. It is possible that the stress condition was not particularly anxiety arousing for most subjects.

Sherman and Jost (1942) in contrast to the above studies did find a number of significant correlations among their physiological variables for neurotic and normal children combined. Although their correlation matrix is based on a mixture of absolute level scores, percent change scores, and scores obtained at different points in a sequence of seven conditions, there does seem to be a cluster of fairly highly intercorrelated variables suggesting some arousal dimension. The measures most highly intercorrelated are hand tremor, percent heart rate change, percent alpha dominance (negative correlations), and respiratory variability. Weybrew (1959) intercorrelated 12 physiological change scores and 4 personality ratings. The physiological measures were obtained before and after the subjects were subjected to a standardized situational stress. Correlations were in

general low among the physiological change scores, and the results of a factor analysis were not easy to interpret.

There are just not enough studies with enough significant correlations between change scores to attempt any generalizations from the results. A general problem encountered in working with autonomic change scores is with respect to the type of transformation, if any, to use. Correlations, for example, among Lacey's (1956) autonomic lability scores would appear to provide a more meaningful picture of the tendency of measures to covary than would be obtained by using absolute change, percentage change, or most other transformations, since as previously mentioned Lacey's score more adequately partials out the usual high negative correlation between change and initial level. The degree to which correlations among autonomic change scores can be affected by partialing out the correlation with initial level is shown in the results of Mandler and Kremen (1958). They intercorrelated scores obtained under a failure stress condition from five different response systems (GSR, heart rate, respiration, face temperature, and blood volume) including in some cases absolute change scores along with Lacey's autonomic lability score. Absolute heart rate change yielded a correlation of .27 with change in respiration rate, whereas heart rate with initial level partialed out yielded a correlation of $-.17$; or in another case absolute heart rate change correlated only .02 with inspiration amplitude (with initial level of inspiration amplitude partialed out) but heart rate with initial level partialed out correlated .31 with the same measure. It is clear that correlations among autonomic measures will be greatly affected by the way in which the relation to initial level is handled.

The findings of Lacey (1950), Lacey and Van Lehn (1952), Lacey, Bateman, and Van Lehn (1953), even though based on stressors that for the most part cannot be accepted as clearly anxiety arousing, provide such a strong argument for individual patterns of autonomic response that they should not be ignored in this context. Using various samples (college students and mothers of children in the Fel's longitudinal research program) and various stressors (cold pressor test, hyperventilation, mental-arithmetic, and word fluency), Lacey et al. (1953) find that different subjects have different patterns of autonomic response which are reproducible over time and are consistent over these different stressors. Thus one subject may respond to the stress by a large increase in heart rate and only a small increase in skin conductance and another may respond with the opposite pattern. To the extent that such findings can be generalized to a clearly fear arousing situation the conclusion is clear that one cannot expect intercorrelations among autonomic change scores to be very substantial. The point to be emphasized here, however, is not that several autonomic measures might not for almost all people increase under anxiety arousing circumstances, but that those measures which show the most or least increase vary from person to person. Such a state of affairs is not necessarily disasterous to one interested in using physiological measures in assessing anxiety. The moral, however, remains clear that for a given individual some physiological measures may be much more sensitive indicators of change in anxiety level than others.

A somewhat similar point of view is espoused by Malmo, Shagass, and Davis (1950) in which they propose the principle of symptom specificity: namely, that psychiatric patients

are inclined to respond to stress of all kinds by a particular physiological mechanism that leads to the particular kind of somatic complaint that the patient may have. Thus, Malmo and Shagass (1949b) found that patients with heart complaints showed greater heart rate and heart rate variability under stress than patients without heart complaints. Specificity of muscle potential reaction was demonstrated by Malmo, Smith, and Kohlmeyer (1956) who showed that for the same patient discussion of hostility conflicts was associated with increased forearm muscle tension and discussion of sex conflict was associated with increased leg muscle tension.

There are other studies in which intercorrelations among a number of physiological measures are reported, such as Wenger (1942, 1948) or Gunderson (1953) in which all measures were obtained under resting conditions. If people manifest varying degrees of an autonomic response pattern determined by the amount of anxiety that they "bring into" the resting situation then such a pattern should show up as a cluster of intercorrelated variables. Wenger's (1942) earlier factor analytic work with children did yield a dimension that he called the autonomic factor, which when unbalanced in the sympathetic direction would appear to be similar to the cluster of autonomic measures associated with experimentally aroused anxiety in the previously described studies. However, in Wenger's (1948) study of aviation cadets, operational fatigue patients, and neurotic patients the case for a clear-cut autonomic factor is shaky. The most striking thing about the reported intercorrelations is their extremely low level. Very few correlations are higher than .15. Gunderson (1953), however, reported intercorrelations among his 12 resting state measures on a subsample of 44 paranoid schizophrenics that were both substantial, for this kind of data, and pervasive. There was a tendency for many of the different autonomic measures to correlate between .20 and .45 in a direction consistent with degree of sympathetic arousal.

In summary, intercorrelations among physiological measures obtained under either resting states or under stress tend to be low and frequently insignificant. There are few studies, however, in which a variety of measures are obtained under a clearly fear arousing situation and where the tendency of change scores to correlate with initial level has been partialed out. Improved measurement technique may also make some of the older studies somewhat obsolete. Nevertheless, the best guess on the basis of present findings is that intercorrelations among physiological measures will be found to be low even with the above-mentioned modifications taken into account. Lacey's work suggests, consistent with the findings of low intercorrelations, that an individual responds to stress with a characteristic pattern of responses. This finding may not be entirely inconsistent with the possibility of there being some pattern of response usually associated with fear. For example, Lacey's findings that subjects showed different response patterns to the stress of doing mental arithmetic may result in part from the fact that some subjects were made angry in the situation and some were made anxious, and that those that were made anxious showed a distinctive pattern from those made angry as Funkenstein et al. found. The chances are that this explanation does not account for all the individual response patterns, and it may be that among subjects made anxious there still remain different response pat-

terns. The meaning of these different response patterns, which could be few in number, may be clarified by further knowledge about their correlation with behavioral and perhaps self-report type measures. It is, of course, possible that future factor analytic or pattern analysis studies will suggest the utility of conceptualizing several different kinds of anxiety states.

Behavioral Measures: Experimental and Group Comparisons

The same question is asked here as was asked with respect to physiological measures; is there some pattern of behavioral effects associated with anxiety that can be distinguished from behavioral effects resulting from other arousal states? The researches most relevant to the question are those in the general area of the effects of stress on performance. These researches, unfortunately, do not provide a clear answer to the question because of two major lacks. First, most such studies tend to be limited to one dependent variable for the good reason that it is much more difficult to measure simultaneously a variety of appropriate behavioral responses than physiological responses. Second, few studies attempt to contrast a fear arousal state with other kinds of arousal states. Another general drawback to most behavioral measures for the purposes of assessment, as will be shown in the studies reviewed, is that their relation to the anxiety construct is not a monotonic one; for example, a low score on a certain performance may be associated with a very low or very high state of anxiety. The studies mentioned below, then, can be seen as only suggestive of measures likely to be especially sensitive to the effects of anxiety, and are not intended to represent an extensive coverage of the research on the effects of stress on performance. Summaries of research in this area are provided by Hanfmann (1950), Lazarus, Deese, and Osler (1952), and more recently Easterbrook (1959).

A loose empirical generalization that emerges from studies in this area is that the kinds of tasks most likely to be affected by stress are learning and memory tasks involving novel or relatively poorly learned responses where incorrect competing responses are both numerous and relatively strong; or perceptual tasks in which conditions are imposed that make appropriate discriminations difficult. Thus, failure stress (usually produced by first ego involving, then failing, and then criticizing the subject) has been shown to impair digit span but not vocabulary items (Moldowsky & Moldowsky, 1952); impair recall of incidental learning but not recall of material explicitly instructed to be learned (Aborn, 1953); and impair relearning of a serial list of nonsense syllables (Smith, 1954). Stress imposed by implying that the subject is neurotic or maladjusted on the basis of projective test responses has been found to impair performance on abstract reasoning, the Holsopple Concept Formation Test, and mirror tracing (Beier, 1951); and to produce more perseveration of incorrect responses on the Luchins Water Jar Task (Cowen, 1952).

A number of studies in which anxiety is introduced by separating subjects into high and low anxiety groups on the basis of the Taylor MAS (1953) provide evidence not only that the detrimental effect of anxiety becomes greater as the strength and number of incorrect competing responses involved in the task increases, but also, for the levels of anxiety involved, that performance is enhanced for the high anxiety subjects on some tasks when the correct response is very dominant. The incorrect competing

responses are usually introduced by increasing the similarity and sometimes also by decreasing the association value of items in a serial learning task, or by both increasing intralist similarity and decreasing similarity between pairs in a paired associate learning task. Lucas (1952), Montague (1953), Farber and Spence (1953), Lazarus, Deese, and Hamilton (1954), Taylor and Chapman (1955), Spence, Farber, and McFann (1956), and Spence, Taylor, and Ketchel (1956), all reported evidence for this relationship. The findings of greater ease of eyeblink conditioning in groups of high anxious as opposed to low anxious subjects (Spence & Farber, 1953; Spence & Taylor, 1951; Taylor, 1951) are also consistent with this general proposition.

Lucas (1952) also studied the effect of experimentally induced failure upon performance as a function of the strength of the incorrect competing responses (manipulated by varying the number of duplications of consonants in a series of consonants being used in an immediate recall task). He found no main effect associated with number of duplications nor any interaction with four degrees of experimentally induced failure. No other studies were found in which anxiety was induced experimentally and its effect upon performance studied where the strength of the incorrect responses was systematically varied within the confines of the same task.

A few studies have made use of real life stress situations that probably meet the need for a really anxiety arousing condition better than the experimental procedures used in the other studies. Beam (1955) obtained measures before doctoral oral examinations and opening night performances in plays as well as at a less stressful period in the subject's life, and found marked impairment in learning a serial list of nonsense syllables, and an increase in palmar sweat and GSR conditioning rate under stress as compared to nonstress. Basowitz et al. (1955) reported a tendency for digit span to be impaired for soldiers undergoing paratroop training as compared to a control group, and Wright (1954) likewise found impairment in digit span in patients confronted with the threat of surgery.

One kind of behavioral measure that would appear promising from an assessment point of view is speech disturbance. Mahl (1956, 1959) has developed a system for reliably scoring speech disturbances of various kinds and has shown certain of these disturbances to be related to variation in anxiety as assessed in psychotherapeutic interviews. Dibner (1956) has employed a similar measure.

In the perceptual area Postman and Bruner (1948) reported impairment in the tachistoscopic perception of three-word sentences under failure stress. Rosenbaum (1953) found greater stimulus generalization under strong shock than weak shock. Smock (1957) reported greater intolerance of ambiguity in a perceptual task under stress than nonstress. Korchin and Basowitz (1954), and Moffitt and Stagner (1956) found increased perceptual closure during paratroop training and experimental threat, respectively.

In studies using group comparisons Angyal (1948) found more impairment in the recognition of patterns of letters under brief exposure conditions in high anxiety patients than other patients. Krugman (1947) and Goldstone (1955) found the threshold for flicker fusion to occur at a lower frequency for anxious than nonanxious subjects.

Eriksen and Wechsler (1955) ingeniously attempted to separate the

effects of anxiety (shock induced) on response processes as opposed to sensory discrimination, and concluded that anxiety results in restricted and stereotyped response preferences but does not impair sensory discrimination.

In the studies reviewed so far in this section the effect of stress has been in general to impair performance. There are many studies, however, in which improved performance is associated with stress. Thus, Steisel and Cohen (1951) and Truax and Martin (1957) found improved performance on simple arithmetic problems as a result of failure stress; and Spence (1957) found better recall of words failed on an anagrams task than words successfully completed.

Likewise studies in which groups have been divided on the basis of self-report measures of general anxiety level also indicate that failure stress can lead to improved performance for some subjects. Thus Lucas (1952), Waterhouse and Child (1954), Williams (1955), and Sarason (1956) found that low anxiety subjects tend to improve under stress and high anxiety subjects tend to show impairment under stress.

Thus, to the extent that failure stress arouses anxiety, this construct appears to be associated with both improvement and impairment of performance. These seemingly contradictory findings are in part reconciled in a study by Stennett (1957), who instead of employing just one stress and one nonstress condition attempted to set up four degrees of intensity of motivation. He found that tracking performance improved at first as the rewards for correct performance increased but then showed impairment under the most extreme condition involving a large bonus for high level performance and threat of electric shock if this level was not reached. He also obtained palmar conductance and muscle potential measures on his subjects and found these measures to increase monotonically as a function of increased "motivation." Several authors, consistent with this study and the others previously described, have proposed that adequacy of performance is an inverted U shaped function of some arousal, activation, or emotional state—for example, Woodworth and Schlosberg (1954), Malmo (1957).

Thus there appear to be two rather loose empirical generalizations that can be reached on the basis of the studies reviewed in this section: (*a*) that tasks involving relatively stronger and more numerous competing responses are more subject to the impairing effects of stress, and (*b*) increasing stress results in improved performance up to a point and impairment thereafter. There is no particular evidence in this area to warrant the separation of anxiety as a construct from other more general constructs such as "arousal," "activation," or "drive."

Somewhat differing theoretical formulations have been proposed to account for the empirical generalizations described. Easterbrook (1959) makes a plausible case for the idea that many of the disorganizing effects of emotion can be accounted for on the basis of cue utilization: namely, that increased "drive" or "emotion" leads to a constriction of the perceptual field or decrease in the number of cues that can be attended to. The Iowa theorists, on the other hand (Spence, 1958), employ the concept of drive and its hypothesized multiplicative relationship to habit strength to account for many of the effects of stress on performance; and Child (1954), Child and Waterhouse (1953), and Sarason, Mandler, and Craighill (1952) emphasize the ir-

relevant competing responses specifically associated with stress on the basis of the past learning.

If anxiety proves to be a distinguishable arousal state, research on its effects on performance would be greatly facilitated if it could be assessed independently, perhaps by physiological measures, from the performance being studied. The utility of this approach is shown in Stennett's study, where it was not necessary to assume that experimental conditions were effective, or to rely upon some paper and pencil measure in determining the presence or magnitude of the motivational or emotional arousal state, but where instead the palmar conductance and muscle potential measures provided more direct evidence of the degree of arousal.

In summary, no studies were discovered in which several objectively measured behavioral characteristics were obtained simultaneously (or almost so) with a variety of physiological measures under conditions likely to be very fear arousing; much less, studies that in addition contrasted different types of arousal states. On the basis of the one and two variable type studies, though, it seems likely that some fairly simple learning, immediate memory, or perceptual tasks could be developed that would be sensitive to changes in anxiety level. It is possible that a few such tasks along with physiological measures could in the future help define more clearly the anxiety response pattern. Although, in general, improved methods of continuous anxiety measurement will probably contribute more to the study of the effects of anxiety on behavior than vice versa.

Behavioral Measures: Intercorrelations

Studies oriented toward assessing the intercorrelations among a number of behavioral manifestations of anxiety are beset by a special problem. Physiological measures can usually be obtained simultaneously but many behavioral effects of anxiety can be assessed only by presenting the subject with a series of tasks to perform. Unknown order effects may well distort the obtained correlations.

There have been several studies of this type in which a number of behavioral measures, selected on the basis of previously reported relationships to anxiety, were intercorrelated. Martin (1958, 1959) in two successive studies using college subjects, found the intercorrelations to be quite low, but a factor analysis still suggested the presence of a dimension that might be labeled anxiety. In the second study some of the measures that had the higher loadings on the anxiety factor were the Taylor MAS, .41; time to learn a complex (five choice) verbal maze, .40; errors in learning of paired associate nonsense syllables with high intralist similarity but low similarity between pairs, .39; tremors on a manual dexterity task, .39; an anxiety check list, .27. A simple verbal maze (two choice) and a paired associate list involving low intralist similarity and high similarity between pairs had zero order loadings on the factor. The loadings with respect to the two kinds of paired associate lists and the two kinds of verbal mazes are consistent with the notion that tasks involving stronger competing responses are more sensitive to the effects of anxiety. A somewhat more prominent factor that also emerged in both studies was interpreted as a motivational factor, that is, a dimension reflecting how hard these college subjects tried on a number of the tasks. Such individual differences in motivation were postulated to be relatively independent of the subjects' anxiety level. A third

factor of some generality was identified as intelligence, and yet another factor was entirely defined by self-report measures of anxiety such as the Taylor MAS. Thus performance on a given task such as learning paired associate nonsense syllables with high intralist similarity under mild stress was found not only to be affected by individual differences in anxiety but also by individual differences in motivation, intelligence, and a factor specific to the type of task. Under these circumstances it is easy to see how anxiety variance could frequently be masked by other factors.

Rosenthal (1955), Cattell and Gruen (1955), and Scheier and Cattell (1958) reported several factor analytic studies in which a variety of self-report, behavioral, and, in some cases, physiological measures were obtained. They found a factor, which they label anxiety, emerging in all their studies that is separable from a number of other personality factors after relatively blind rotations to oblique simple structure. The above studies employed substantial *N*s in five different samples of subjects involving USAF pilot trainees, children, and college students. Upon inspection of the factor loadings on the anxiety factor in these various studies as summarized by Cattell and Scheier (1958b) it becomes apparent, however, that the only measures with high loadings and the only measures whose loadings are consistent from study to study are those based on self-report type measures. Few if any behavioral-physiological measures have loadings over .30 and none of those that do are substantiated in any of the other samples. For example, in Rosenthal's study (1955) the three highest loadings on the anxiety factor were Taylor MAS, .85; questionnaire measure of anxious insecurity, .84; and a questionnaire measure of nervous tension, .70. The other four measures with loadings above .30 were also self-report type measures. Rosenthal obtained several physiological measures under various conditions (GSR, heart rate, salivary volume, systolic blood pressure) and none of these were related to this anxiety factor to any degree. Under these circumstances it does not seem reasonable to accept this factor as necessarily assessing the hypothetical anxiety reaction as formulated in this paper.

Cattell and Scheier (1958b) distinguish between the "trait" of anxiety, inferred from factor analysis of a cross section of measures obtained only once on each subject, and the "state" of anxiety inferred from a factor analysis of change scores from one testing time to another. Correlating change scores in this way is referred to as incremental R technique, and Cattell and Scheier (1958a) report in detail the results of such a study. An interesting innovation in this study, too involved to go into in this paper, was the introduction of different "treatment" conditions into a correlational study, so that it was possible to see, for example, how imminence of academic examinations correlated with the other variables. One of the resulting 14 factors was identified as the "state" anxiety factor and appears to represent an arousal state more closely related to the present theoretical view of anxiety than the previously found trait factor. The self-report measures did not dominate the loadings so much, although the two highest loadings were self-report measures involving an anxiety-tension check list, .41, and a questionnaire scale of tension, .40. In addition though, systolic blood pressure had a loading of .30 and palmar con-

ductance of .26. Perhaps inconsistent with this was the positive loading of volume of saliva, .27. The imminence of an academic examination was negatively loaded, −.25, suggesting that just before an examination the usually anxious person becomes less anxious. The authors propose that "a person beset by vague fears and anxieties loses these anxieties for a while when a real danger threatens."

Holtzman and Bitterman (1956) intercorrelated 41 measures obtained on 135 cadets in an Air ROTC unit. These measures included ratings, personality tests, stress tests, perceptual tests, GSR conditioning, and amount of uric acid and glycine in the urine. The intercorrelations among the different kinds of measures were quite low and a factor analysis yielded seven factors which were almost entirely determined by clusters of measures taken from the same test situation.

There are some important limitations to the factor analytic approach to the study of anxiety. For example, there is no convincing logic to the supposition that simple structure, oblique or orthogonal, yields the most psychologically meaningful dimensions; although intuitively it would seem that some kind of oblique solution would be more meaningful for separating out a cluster of physiological-behavioral measures to be idenified as anxiety as opposed to clusters of measures representing other arousal states, since in all likelihood these various arousal states will be correlated. With respect to rotations in factor analytic studies perhaps it would be better if such rotations were not done blindly but with full knowledge of the nature of the measures, and the final rotation considered frankly for what it is, a post hoc hypothesis about the nature of the dimensions revealed. Confirmation of the interpretation of a given factor and further elucidation of the construct validity (Cronbach & Meehl, 1955) of the assessment procedures can then be ascertained by introducing the factor as a variable in experimental research.

Certainly the selection of measures to be intercorrelated affects the definition of the resulting factors. For example, it may be that in the Cattell studies just described, with the exception of the incremental R technique study, that the high intercorrelations among the self-report measures, which almost entirely define the anxiety factor, are due in part to correlated nonanxiety variance. It is also possible that many of the measures used in the factor analytic studies involve characteristic ways of controlling or reducing anxiety rather than more direct manifestations of the anxiety itself. The Holtzman and Bitterman study serves to point up the fact that in an area where correlations between measures obtained from different response systems are going to be low at best, including clusters of highly intercorrelated measures from the same response system or test situation will inevitably result in factors representing these clusters, at least when the common criteria for simple structure are employed. It is possible that a factor analysis done under such conditions might serve to actually hide some real generalities of response, although there is no indication that such was the case in the Holtzman and Bitterman study.

One cannot conclude on the basis of the researches reviewed in this paper, despite many suggestive leads, that any clear-cut pattern of physiological-behavioral responses associated with anxiety arousal, distinguishable from other arousal patterns has been demonstrated. The status

of anxiety assessment procedures, both in terms of experimental and correlational findings might be clarified by combining some of the best features of the researches described. First one might attempt to measure simultaneously, or nearly so, an extensive battery of physiological measures and a few selected behavioral measures at a time when the subject is relaxed. This would necessitate a preliminary adaptation-to-the-apparatus session. Then the subjects could be tested again under definitely anxiety arousing circumstances, the more realistic the better. A study of the change score patterns and intercorrelations, after correcting where necessary for correlation with relaxed session levels, should provide evidence for an anxiety pattern if it exists. It would then be further necessary to demonstrate that the pattern of responses was distinguishable from patterns associated with other arousal states such as general activation, anger, or sex; otherwise there is no utility in having a construct of anxiety separate from these others.

When more is known about the physiological-behavioral response pattern associated with anxiety, then self-report scales can be constructed which will predict this response pattern in various situations.

REFERENCES

ABORN, M. The influence of experimentally induced failure on the retention of material acquired through set and incidental learning. *J. exp. Psychol.*, 1953, **45**, 225–231.

ANGYAL, ALICE F. The diagnosis of neurotic traits by means of a new perceptual test. *J. Psychol.*, 1948, **25**, 105–135.

AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, **15**, 433–442.

BARCROFT, H., & KONZETT, H. On the actions of noradrenaline and isopropyl noradrenaline on the arterial blood pressure, heart rate and muscle blood flow in man. *J. Physiol.*, 1949, **110**, 194–204.

BASOWITZ, H., PERSKY, H., KORCHIN, S. J., & GRINKER, R. R. *Anxiety and stress.* New York: McGraw-Hill, 1955.

BEAM, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–551.

BEIER, E. G. The effect of induced anxiety on flexibility of intellectual functioning. *Psychol. Monogr.*, 1951, **65**(9, Whole No. 326).

BITTERMAN, M., & HOLTZMAN, W. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.

CATTELL, R. B., & GRUEN, W. The primary personality factors in 11-year-old children by objective tests. *J. Pers.*, 1955, **23**, 460–478.

CATTELL, R. B., & SCHEIER, I. H. Factors in personality change: A discussion of the condition-response incremental design and application to 69 personality response measures and three stimulus conditions. (Advanced Publication No. 9) Urbana, Ill.: Laboratory of Personality Assessment and Group Behavior, University of Illinois, 1958. (a)

CATTELL, R. B., & SCHEIER, I. H. The nature of anxiety: A review of thirteen multivariate analyses comprising 814 variables. *Psychol. Rep.*, 1958, **4**, 351–388. (b)

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

CHILD, I. L., & WATERHOUSE, I. K. Frustration and the quality of performance: II. A theoretical statement. *Psychol. Rev.*, 1953, **60**, 127–139.

CLEMENS, T. L. Autonomic nervous system responses related to the Funkenstein test: I. To epinephrine. *Psychosom. Med.*, 1957, **19**, 267–274.

COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512–519.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

DAVIS, R. C. Response patterns. *Trans. NY Acad Sci., Ser.* II, 1957, **19**, 731–739.

DAVIS, R. C., & BUCHWALD, A. M. An exploration of somatic response patterns: Stimulus and sex differences. *J. comp. physiol. Psychol.*, 1957, **50**, 44–52.

DAVIS, R. C., BUCHWALD, A. M., & FRANK-

MAN, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychol. Monogr.*, 1955, **69**(20, Whole No. 405).

DELARGEY, C., GREENFIELD, A. D. M., MCCORRY, R. L., WHELAN, R. F. The effects of intravenous infusion of mixtures of L-adrenaline and L-noradrenaline on the human subject. *Clin. Sci.*, 1950, **9**, 71–78.

DIBNER, A. S. Cue-counting: A measure of anxiety in interviews. *J. consult. Psychol.*, 1956, **20**, 475–478.

DIMASCIO, A., BOYD, R. W., & GREENBLATT, M. Physiological correlates of tension and antagonism during psychotherapy. *Psychosom. Med.*, 1957, **19**, 99–104.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66**, 183–201.

ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 458–463.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. A test which predicts the clinical effects of electric shock treatment on schizophrenic patients. *Amer. J. Psychiat.*, 1950, **106**, 889–901.

FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Autonomic changes paralleling psychologic changes in mentally ill patients. *J. nerv. ment. Dis.*, 1951, **114**, 1–18.

FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Nor-epinephrine like and epinephrine like substances in psychotic and psychoneurotic patients. *Amer. J. Psychiat.*, 1952, **108**, 652–662.

FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. E. *Mastery of stress.* Cambridge: Harvard Univer. Press, 1957.

GOLDENBERG, M. Adrenal medullary function. *Amer. J. Med.*, 1951, **10**, 622–641.

GOLDENBERG, M., PINES, K. L., BALDWIN, E. DE F., GREENE, D. G., & ROH, C. E. The hemodynamic response of man to norepinephrine and epinephrine and its relation to the problems of hypertension. *Amer. J. Med.*, 1948, **5**, 792–799.

GOLDSTONE, S. Flicker fusion measurements and anxiety level. *J. exp. Psychol.*, 1955, **49**, 200–202.

GUNDERSON, E. K. Autonomic balance in schizophrenia. Unpublished doctoral dissertation, University of California, Los Angeles, 1953.

HANFMANN, E. Psychological approaches to the study of anxiety. In P. H. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 51–69.

HICKHAM, J. B., CARGILL, W. H., & GOLDEN, G. Cardiovascular reactions to emotional stimuli: Effect on cardiac output, A-V oxygen difference, arterial pressure and peripheral resistance. *J. clin. Invest.*, 1948, **27**, 290–298.

HOLTZMAN, W. H., & BITTERMAN, M. E. A factorial study of adjustment to stress. *J. abnorm. soc. Psychol.*, 1956, **52**, 179–185.

JURKO, M., JOST, H., & HILL, T. S. Pathology of the energy system: An experimental-clinical study of physiological adaptive capacities in a non-patient, a psychoneurotic, and an early paranoid schizophrenic group. *J. Psychol.*, 1952, **33**, 183–198.

KORCHIN, S. J., & BASOWITZ, H. Perceptual adequacy in a life stress. *J. Psychol.*, 1954, **38**, 495–502.

KRUGMAN, M. Flicker fusion frequency as a function of anxiety reaction: An exploratory study. *Psychosom. Med.*, 1947, **9**, 269–272.

LACEY, J. I. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, **43**, 338–350.

LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. NY Acad. Sci.*, 1956, **67**, 123–164.

LACEY, J. I., BATEMAN, DOROTHY E., & VAN LEHN, RUTH. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15**, 8–21.

LACEY, J. I., & VAN LEHN, RUTH. Differential emphasis in somatic response to stress: An experimental study. *Psychosom. Med.*, 1952, **14**, 71–81.

LAZARUS, R. S., DEESE, J., & HAMILTON, R. Anxiety and stress in learning: The role of intraserial duplication. *J. exp. Psychol.*, 1954, **47**, 111–114.

LAZAZUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

LEWINSOHN, P. M. Some individual differences in physiological reactivity to stress. *J. comp. physiol. Psychol.*, 1956, **49**, 271–277.

LUCAS, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.

MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–15.

MAHL, G. F. Measuring the patient's anxiety during interviews from "expressive" as-

pects of his speech. *Trans. NY Acad. Sci.*, 1959, **21**, 249–257.

MALMO, R. B. Experimental studies of mental patients under stress. In M. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MALMO, R. B., BOAG, T. J., & SMITH, A. A. Physiological study of personal interaction. *Psychosom. Med.*, 1957, **19**, 105–119.

MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24. (a)

MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, **11**, 25–29. (b)

MALMO, R. B., SHAGASS, C., BELANGER, D. J., & SMITH, A. A. Motor control in psychiatric patients under experimental stress. *J. abnorm. soc. Psychol.*, 1951, **46**, 539–547.

MALMO, R. B., SHAGASS, C., & DAVIS, J. F. A method for the investigation of somatic response mechanisms in psychoneurosis. *Science*, 1950, **112**, 325–328.

MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular tension in psychiatric patients under stress. *J. clin. exp. Psychopathol.*, 1951, **12**, 45–66.

MALMO, R. B., & SMITH, A. A. Forehead tension and motor irregularities in psychoneurotic patients under stress. *J. Pers.*, 1955, **23**, 391–406.

MALMO, R. B., SMITH, A. A., & KOHLMEYER, W. A. Motor manifestiation of conflict in interview: A case study. *J. abnorm. soc. Psychol.*, 1956, **52**, 268–271.

MANDLER, G., & KREMEN, I. Autonomic feedback: A correlational study. *J. Pers.*, 1958, **26**, 388–399.

MARTIN, B. A factor analytic study of anxiety. *J. clin. Psychol.*, 1958, **14**, 133–138.

MARTIN, B. The measurement of anxiety. *J. gen. Psychol.*, 1959, **61**, 189–203.

MILLER, N. F. Central stimulation and other new approaches to motivation and reward. *Amer. Psychologist*, 1958, **13**, 100–108.

MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354–357.

MOLDAWSKY, S., & MOLDAWSKY, PATRICA C. Digit span as an anxiety indicator. *J. consult. Psychol.*, 1952, **16**, 115–118.

MONTAGUE, E. K., The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.

POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–324.

ROSENBAUM, G. Stimulus generalization as a function of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.

ROSENTHAL, IRENE. A factor analysis of anxiety variables. Unpublished doctoral dissertation, University of Illinois, 1955.

SARASON, I. G. Effects of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–259.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SCHACHTER, J. Pain, fear and anger in hypertensives and normotensives. *Psychosom. Med.*, 1957, **19**, 17–29.

SCHEIER, I. H., & CATTELL, R. B. Confirmation of objective test factors and assessment of their relation to questionnaire factors: A factor analysis of 113 rating, questionnaire, and objective test measurements of personality. *J. ment. Sci.*, 1958, **104**, 608–624.

SCHIFF, E., DOUGAN, C., & WELCH, L. The conditioned PGR and the EEG as indicators of anxiety. *J. abnorm. soc. Psychol.*, 1949, **44**, 549–552.

SELYE, H. *Stress*. Montreal: Acta, 1950.

SHAGASS, C. The sedation threshold: A method for estimating tension in psychiatric patients. *EEG clin. Neurophysiol.*, 1954, **6**, 221–233.

SHAGASS, C. Anxiety, depression, and the photically driven electroencephalogram. *AMA Arch. Neurol. Psychiat.*, 1955, **74**, 3–10. (a)

SHAGASS, C. Differentiation between anxiety and depression by the photically activated electroencephalogram. *Amer. J. Psychiat.*, 1955, **112**, 41–46. (b)

SHAGASS, C., & NAIMAN, J. The sedation threshold, manifest anxiety, and some aspects of ego function. *AMA Arch. Neurol. Psychiat.*, 1955, **74**, 397–406.

SHERMAN, M., & JOST, H. Frustration reactions of normal and neurotic persons. *J. Psychol.*, 1942, **13**, 3–19.

SMITH, J. G. Influence of failure, expressed hostility, and stimulus characteristics on verbal learning and recognition. *J. Pers.*, 1954, **22**, 475–493.

SMOCK, C. D. Recall of interrupted and noninterrupted tasks as a function of experimentally induced anxiety and motivational relevance of the task stimuli. *J. Pers.*, 1957, **25**, 589–599.

SPENCE, D. P. Success, failure, and recognition threshold. *J. Pers.*, 1957, **25**, 712–720.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the US as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., TAYLOR, J., & KETCHEL, RHODA. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

STEISEL, I. M., & COHEN, B. D. The effects of two degrees of failure on level of aspiration and performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 79–82.

STENNETT, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG clin. Neurophysiol.*, 1957, **9**, 131–138.

SWAN, H. J. C. Noradrenaline, adrenaline, and the human circulation. *Brit. med. J.*, 1952, **1**, 1003–1006.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A., A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & CHAPMAN, JEAN P. Anxiety and the learning of paired-associates. *Amer. Psychol.*, 1955, **68**, 671.

TERRY, R. A. Autonomic balance and temperament. *J. comp. physiol. Psychol.*, 1953, **46**, 454–460.

TRUAX, C. B., & MARTIN, B. The immediate and delayed effect of failure as a function of task complexity and personalization of failure. *J. abnorm. soc. Psychol.*, 1957, **55**, 16–20.

ULETT, G. A., GLESER, G., WINOKUR, G., & LAWLER, A. The EEG and reaction to photic stimulation as an index of anxiety-proneness. *EEG clin. Neurophysiol.*, 1953, 23–32.

WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1954, **21**, 298–311.

WELSH, L., & KUBIS, J. F. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.

WENGER, M. A. A study of physiological factors: The autonomic nervous system and the skeletal musculature. *Hum. Biol.*, 1942, **14**, 69–84.

WENGER, M. A. Studies of autonomic balance in Army Air Forces personnel. *Comp. psychol. Monogr.*, 1948, No. 101.

WENGER, M. A. Pattern analyses of autonomic variables during rest. *Psychosom. Med.*, 1957, **19**, 240–244.

WEYBREW, B. B. Patterns of reaction to stress as revealed by a factor analysis of autonomic-change measures and behavioral observation. *J. gen. Psychol.*, 1959, **60**, 253–264.

WHITE, B. V., & GILDEA, R. F. "Cold pressor test" in tension and anxiety: A cardiochronographic study. *Arch. Neurol. Psychiat., Chicago*, 1937, **38**, 964–984.

WILLIAMS, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.*, 1955, **51**, 573–580.

WISHNER, J. Neurosis and tension: An exploratory study of the relation of physiological and Rorschach measures. *J. abnorm. soc. Psychol.*, 1953, **2**, 253–260

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

WRIGHT, M. W. A study of anxiety in a general hospital setting. *Canad. J. Psychol.*, 1954, **8**, 195–203.

(Received February 10, 1960)

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN